The (Bene)fits of Compiling a
Specialized Database

by Anne Wade

Abstract

The Centre for the Study of Classroom Processes (CSCP) at Concordia University in Montreal recently completed a three year project to compile a specialized database on the successful pedagogical approach of cooperative learning. Using the BiblioLinks software, the results of CD-ROM searches executed on a variety of source databases, were transferred into an in-house database. Utilizing the ProCite bibliographic management software, a package designed to store bibliographic information, the database was customized according to the needs of the CSCP. The benefits of establishing a specialized database included the ability to perform extensive searches on multidisciplinary sub-topics within the area of cooperative learning. However a number of problems occurred throughout the course of compiling the database, the majority of which related to limitations with the hardware and software. This article provides a chronology of those problems. Issues related to copyright and the publication of this database in the form of a comprehensive bibliography, are also addressed.

The Centre for the Study of Classroom Processes (CSCP), is a research centre affiliated with the Department of Education at Concordia University, Montreal, devoted to the study of effective pedagogical techniques. Researchers at the CSCP have studied such topics as group composition (Abrami et al, in progress), dramatic play in early childhood education (Howe, et al., in press), mental modeling (Schmid et al, in progress), the effectiveness of student ratings (Abrami et al, 1991), group size and microcomputer learning (Baron & Abrami, 1992) and cooperative learning strategies (Abrami et al., under review; Abrami et al., 1993). It is this latter subject area in which much of the CSCP research has been performed over the years. Because of this, researchers at the CSCP felt that it would be beneficial and cost-effective to establish a specialized in-house database of bibliographic information pertaining to cooperative learning and related areas. Bibliographic records would be retrieved from a variety of subject databases using the available online and CD-Rom systems and subsequently transferred into one main database. This database could then be used by CSCP researchers for in-house searches on sub-topics within cooperative learning. Bibliographies would be generated quickly and easily for the articles published at the CSCP and the database could also act as an online catalogue for the material housed within the CSCP Resource Centre.

In line with the above, Williams (1986) discusses the benefits of creating speciality databases due to the ease and availability of inexpensive existing information technology and suggests:

> The ultimate user...is the customer of the user who has now become a producer of a speciality database. There is added value by virtue of some added processing, screening, manipulation, reworking of data to make the information more palatable to a given market.

However, what was thought would take about six months to complete, actually took over three years and the concentrated effort of the part-time manager of the Resource Centre at the CSCP. Changes in the management, which occurred three times throughout the implementation of the project, contributed to the length of the project. What follows is a synopsis of some of the problems experienced and the issues involved in publishing the results of these efforts. These include: problems related to software selection, complexity of search strategies, limitations of both the hardware and software, lack of standardization between the CD-ROM databases, and copyright restrictions.

Selecting the Software

To begin with a package had to be selected which would meet the requirements of this project. Knowing that the literature searches would retrieve thousands of citations due to the vast quantity of information being published in the area of cooperative learning, the software had to allow for the creation of a large database; it had to provide for the creation of user-defined workforms to input not only standard bibliographic information but also various unique fields specific to the CSCP; and it had to have the capability of transferring machine-readable records from a variety of online and CD-ROM databases. Two popular bibliographic management packages were compared: the ProCite Bibliographic System offered by Personal Bibliographic Software Inc. (PBS), and Reference Manager by Research Information Systems. After careful consideration, ProCite (Version 1.4) was purchased as it was geared more towards library use. Various features included twenty-four "ready to use" or default workforms, each of which consisted of bibliographic fields specific to a particular type of material. For example, a workform existed for each of the following: journals, books, non-print

material, reports, and dissertations (see Table I). ProCite also allowed the user to create up to six "user-defined" or customized workforms for those who wished to modify or add additional bibliographic fields and therefore did not want to use the default forms. Other noteworthy features included the ability to import records from remote databases (although a separate transferring package would have to be purchased for this), and its virtual unlimited record and database size. Conversely, Reference Manager only provided workforms for book, journals and unpublished works with no option to input new fields or to create your own workform. It seemed to be targeted more for the researcher. Because the CSCP also wanted to include its holdings, which consisted of many different types of material, ProCite seemed the logical choice.

Two additional purchases were necessary before the project could get underway. A subscription to the online vendor Bibliographic Retrieval Service (BRS) provided access to a number of relevant online bibliographic databases such as ERIC, Dissertation Abstracts International and PsycInfo. Finally, the purchase of the BiblioLinks software for BRS (also from PBS) was necessary as it formatted machine readable records retrieved from an online search into a format recognized by ProCite. Following these purchases, the information professional who had started the project, moved away and a second professional, who had experience using ProCite, was hired.

Formulating the Search Strategies

Effective search strategies had to then be devised for each of the subject databases. It was critical that the strategies be broad enough to retrieve all relevant material with the smallest number of false hits. After numerous meetings with the researchers, a chart was devised which listed both relevant and unwanted terms and various sample searches were performed. Eventually a comprehensive search strategy was formulated which consisted of over 100 search statements tailored for the ERIC database that was offered online via BRS. Less extensive strategies were then devised for PsycInfo, Sociofile, Dissertation Abstracts International (DAI), Education Library, Canadian Education Index, and the Ontario Education Resources Information System (ONTERIS) databases.
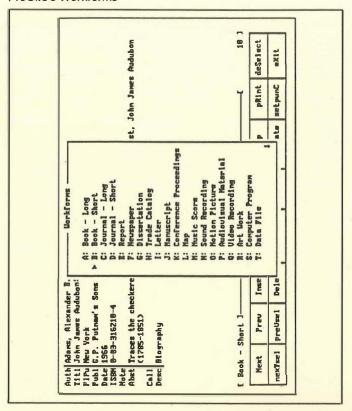
A ProCite database was created and customized workforms (or screens) were designed for each of the following types of bibliographic material: Books, journals, dissertations, reports, conference proceedings and non-print material. Fields were adapted within each of the workforms based on the projected needs of the CSCP. For example, fields were added to indicate: whether the item was located at the CSCP; ordering information; author affiliation and address; and other relevant information.

The Advent of CD-ROM Technology

Six months into the project there was, once again, a change in management and this author was hired to take over. CD-ROM technology had just arrived at Concordia, and three of the main

Table 1.
ProCite's Workforms

source databases (ERIC, PsycLit, and SocioFile) that were to be used on this project were purchased by the university from SilverPlatter. Because searching on CD-ROM was less expensive than online (Concordia charged $5/hour vs BRS's connect time charges) it was decided that CD-ROM searches would be performed whenever possible. Thus a second transferring package, BiblioLinks for SilverPlatter, was purchased from PBS.

The next couple of weeks involved testing the CD-ROM hardware and adapting the search strategies to the SilverPlatter software. When it was time to run the searches, initial problems related to both hardware and software constraints on the CD-ROM systems: SilverPlatter would not accept more than 100 search statements; search strategies could not be saved for quarterly updates of search results; and the complexity of the strategies — long search statements and the combination of numerous terms (e.g., *SMALL near1 (TEAM* OR GROUP*) near2 (LEARN* or INSTRUCTION or WORK*)* would jam the hardware. This was particularly frustrating as the computer would seize up, usually around the 90th search statement, thus requiring the re-input of the entire search strategy. The search strategy could not be sectioned into smaller units as the statements were all interrelated and had to be executed in one session. Also, because the retrieved sets were so large, the computer would jam after having downloaded 590 of the 600 records retrieved, with the only option being to reboot the computer and re-enter the entire search session. The majority of these problems occurred as SilverPlatter was refining its software and Concordia

was upgrading its hardware. Thus, the project had to be put on hold a number of times in order to wait for the respective hardware/software upgrades. Months later the searches were implemented successfully.

BiblioLinks and the Transferring of Records into ProCite

The results of all the CD-ROM searches were saved as ASCII files on floppy disks. Before transferring the records into a ProCite file it was necessary to retrieve the ASCII files in WordPerfect for what should have been minimal editing (e.g., adding a notation at the beginning of the file stating the name of the database records were retrieved from). When transferring bibliographic records into a database management package, it is critical that the fields correspond so that the information is sent into the appropriate fields. This can be tricky as the source databases, such as ERIC and PsycLit, all contain different fields and/or field labels; some abbreviate the field label, while others do not. As well, each database uses a different bibliographic style. For example, one index may use the format "Online, v33, no.1, 1992" while another uses "Online, vol. 33(1), May 1993". Even within the same database the source field can change. For instance, PsycLit records changed their format in 1977. The advantage of a transferring package such as BiblioLinks is that it is designed to recognize such differences and little, if any, editing on the ASCII files should be necessary.

BiblioLinks offered separate software packages for each of the various online systems. However, due to the fact that CD-ROM technology was so new, PBS did not have packages for many of the different CD-ROM systems such as those offered by, University Microform International (UMI), Micromedia and the Institute of Scientific Information (ISI). Therefore, for those CD-ROM searches performed on UMI's Dissertation Abstracts International and Micromedia's Canadian Education Index, macros had to be developed within WordPerfect and subsequently run on the ASCII files in an attempt to adapt the records into a format similar to that of SilverPlatter. Various field labels had to be abbreviated, renamed, and re-assigned for each of the different databases in order for them to be recognized by the SilverPlatter BiblioLink.

However, even the ASCII files from the SilverPlatter searches had to be edited for successful data transfer and fields had to be re-assigned within BiblioLinks (fortunately a relatively straight-forward task). Some of this reformatting was due to the fact that user-defined workforms had been developed within ProCite. In retrospect, it would have saved a lot of time if ProCite's default workforms had been used and perhaps adapted to the needs of the CSCP.

Once the ASCII files were formatted appropriately, they were ready to be uploaded into the ProCite database. However, there was more editing required within ProCite! Because ProCite's Version 1.4 did not allow for global changes, each record had to be edited individually. For example, journal fields were not always parsed properly, and information had to be moved into the appropriate fields. Also, hard carriage returns that were inserted

at the end of each line in the abstract field had to be deleted. ProCite has some word processing features, such as Block and Move, thus making editing a little easier, but it was still time-consuming.

One major problem occurred with the ERIC database, which indexes journals, reports, conference proceedings, dissertations and books. Each different type of bibliographic material should be transferred into the corresponding workform (e.g., journal articles should transfer into the Journal workform, reports into the Report workform etc.). However, BiblioLinks is not able to distinguish between different bibliographic records within the same ASCII file and all records were dumped into the Journals workform. Within BiblioLinks there is a feature which can be changed prior to transferring the records. This feature allows one to re-assign a database, such as ERIC, so that records are transferred into a workform other than the default. What this meant, however, was that individual macros had to be set up within WordPerfect and run in the ERIC ASCII file to first move all the books, then the reports, then the conference proceedings and finally the dissertations, into their own respective ASCII files. Each file was then subsequently transferred into ProCite once the BiblioLinks command had been changed to reflect the type of bibliographic material that would be transferred from the ASCII file. Unfortunately the macros were not fool-proof and it was easier to have the journals in one file and all the other material in another. BiblioLinks was customized to transfer records in the "other material" file into the Reports workform. Workforms for records which were not reports were then changed to the appropriate workform within ProCite. Once the workform was changed, some information had to be moved into the appropriate field.

It had been decided that each record would be assigned a code which noted the database from which it had been retrieved. For example, ERIC records were assigned CL1, PsycLit records CL2, SocioFile records CL3 etc. These codes had to be assigned to each record individually as there was no feature in Version 1.4 which globally copied information to a field. In addition, the duplicate detection command was run after records were transferred from each database to eliminate all duplicate records. It had been determined that priority would be given to ERIC records and next to PsycLit records as the descriptors assigned to ERIC records would be more relevant. If a record was picked up on two or more searches, it was coded to reflect this.

Another task was to flag those records which could be found at the CSCP, because the CSCP has an extensive collection of journal articles which has been collected over the years for various research projects. It was expected that all the articles on cooperative learning that were housed at the CSCP would be retrieved on the searches and it would be useful to flag these. A Research Assistant was hired to input into a ProCite file, the author and title of all the articles housed at the CSCP. This file was then merged with the main database and the duplicate detection feature was subsequently run. All those records which were retrieved on a search and located at the CSCP were then coded to reflect this fact and the skeleton record was deleted. Any remaining articles that were at the CSCP but had not been picked up on a search were subsequently edited and the remaining bibliographic information was manually entered.

PBS released ProCite's Version 2.0 during the editing phase of the project and it offered both a Global Search and Replace and a Global Copying command which saved considerable time. But even with the improvements in Version 2.0, some of the changes, most of which related to capitalization, had to be done one record at a time. For example, contrary to the American Psychological Association (APA) standards, ERIC capitalizes the first character of each word in a title; some of the databases use full capitalization for index terms, others do not; and UMI's DAI uses full capitalization for their entire record including the abstract. It should also be mentioned that because PBS did not release an upgrade to BiblioLinks (Ver.1.1) along with their upgrade to ProCite, ASCII files had to be transferred into a database using the old ProCite (Ver.1.4) and then converted to the new version of ProCite.

Downloading the Results into WordPerfect

After many months of time consuming editing, and the need to upgrade the hardware to a 486, 50Mhz computer as the database was growing at an alarming rate, the results were beginning to reach fruition. Subject searches were run for researchers and graduate students at the CSCP. They were sufficiently impressed and the results of these searches were downloaded into WordPerfect to be printed as subject bibliographies. Still more editing required within WordPerfect. Because ProCite does not move the title into the author field for anonymous articles, the date had to be moved. Also bold and underlined characters would not transfer into the WordPerfect file, so additional macros had to be set up. Once again, some of this editing was reduced with Version 2.0 as it provided the ability to save files in WordPerfect format and ASCII characters were downloaded correctly. Needless to say, a lot of time would have been saved if this new version had been available at the start of the project.

Sharing the Results

After three years, the CSCP's specialized database on cooperative learning was finally up and running. Numerous printed sources, including: the Education Index, British Education Index, Educational Technology Index, Books in Print, publisher's catalogues, and various directories to audio-visual material were scanned and all relevant material was manually entered into the database. The final product consisted of over 10,000 citations to books, journal articles, dissertations, reports, conference proceedings, curriculum and non-print material. Each record consisted of a full citation with abstract, keywords and the availability of the document (when provided). Updates of the manual and CD-ROM searches were implemented as the sources were updated to ensure currency of the database. Numerous in-house searches were performed relating to specific areas within cooperative/collaborative learning, such as: multiethnic students, early childhood education, second language instruction, computer-assisted instruction and disabled students. With Version 2.0, global changes could be made to selected records or to the entire

database; ProCite search strategies could be saved directly within the program; even the search results could be saved; and comprehensive searches could be run relatively quickly with the Quick Search feature. Results were so satisfying that it was decided to share them with others outside the university. There were two options — the CSCP could publish a bibliography on cooperative learning in printed format, or the database could be offered to the public in machine-readable form (on diskette or CD-ROM). It was decided that both options would be pursued as different markets would be targeted for each of the two products.

Having downloaded such a vast quantity of information, it was unclear as to whether the CSCP would be within the copyright boundaries in publishing the results, especially due to the fact that the abstracts were copyrighted. The end-user is unaware of how one can use the information retrieved on a CD-ROM search. This is due to the fact that they do not know the terms within the licensing agreement between the institution, in this case, Concordia libraries, and a database supplier, such as the American Psychological Association. This is a common problem within libraries and Duggan (1991) suggests that signs should be posted next to the CD-ROM workstations, much like those next to the photocopy machines, which state the copyright restrictions as outlined in the licensing agreements. This is an excellent idea but one that is not being practised in most places at the moment.

Because of this uncertainty, requests had to be sent out to all the various database suppliers — could the specialized database be published in printed form? "No problem," said the Educational Resources Information Center (ERIC), "bibliographic data has been generated with the support of tax dollars and is in the public domain and may be used without restriction." In fact, ERIC's objective is to achieve widespread dissemination of its data at the lowest possible cost and they encourage projects such as this. University Microfilms Inc. said there was no problem as long as it was stated that the full text of the DAI records were available from UMI. The American Psychological Association was not quite as forthcoming. They required written authorization for replication of PsycLit records in a printed bibliography. Similarly, the International Sociological Association would give permission only after receiving a complete list of accession numbers for the Sociofile records that would be used.

It was felt that the above requirements could be met and samples of the database were then sent out to a variety of American, Canadian and British publishers in an attempt to find a publisher for the bibliography. The general consensus was that bibliographic information lends itself to electronic publishing and the printed format was passé. Only one publisher expressed an interest. However, they required camera-ready copy and would provide no funding to pay for this. They also required that the CSCP would agree to purchase 100 copies of the final product.

The idea of a printed comprehensive bibliography on cooperative learning began to fade. Efforts were then focused on finding a way to offer the database in machine-readable form. Fortunately, the Ontario Institute for Studies in Education (OISE) conveyed an interest in offering our database on CD-ROM along with their ONTERIS database. The database suppliers were then approached once again to find out their copyright restrictions for

publishing the database in electronic form. The copyright boundary had been hit. Although both ERIC and UMI were agreeable, no other commercial database supplier would give permission to publish these results in electronic form. Interestingly, UMI requested a letter which outlined the entire project. This was reviewed by a committee which reached an affirmative decision weeks later. Our request was obviously scrutinized and given careful consideration. UMI must have realized that our desire to publish a bibliography in electronic form represented the tip of the iceberg in what would seem to be a rapidly expanding electronic revolution.

What Does the Future Hold
for Specialized Databases?

Having managed the last two years of this project only to find at the very end that it would be difficult to share the results with the public, was extremely disappointing. But the quest continues. The CSCP has recently approached various funding agencies in an attempt to obtain a publications grant. Provided one of these proposals is approved, the publisher who expressed an interest, but required camera-ready copy, will publish the comprehensive bibliography in the print format. In the meantime, CSCP members, educators, researchers and graduate students have certainly benefitted from the numerous in-house searches which have been performed and which will continue to be performed into the future. Additionally, the CSCP has developed a partnership with the International Association for the Study of Cooperation in Education (IASCE) in which customized bibliographies are offered to IASCE members — this is allowed as long as it is a non-profit venture. In fact the CSCP is presently considering compiling another specialized database on post-secondary education.

Many of the problems that were experienced throughout this project would not exist to the same extent now due to advances in the software — bibliographic management systems, transferring packages (e.g., PBS is presently working on an upgrade for BiblioLinks), and CD-ROM systems have all been improved over the last couple of years. In May, PBS came out with ProCite Version 2.1 which corrected ProCite's serious problem of not recognizing extended memory. Because of this problem, the user would receive an error message when trying to display "pop-up menus" and authority lists within a large database, even though sufficient extended memory was present. Other problems would occur when editing or deleting a record — ProCite would prompt you to save the record immediately as the computer was running out of memory. With Version 2.1, ProCite now recognizes up to 4Mb of available *expanded* memory — but due to ProCite's programming language it is still not able to recognize *extended* memory. Version 2.1 also fine-tuned other less critical tasks such as the Repair, Quick Search, and Duplicate Detection commands. It is expected that PBS will continue to provide improvements to its software and will develop additional BiblioLinks packages for all the various CD-ROM systems.

More significantly, we are beginning to see systems such as Innovative Interface's InnoPac provide the option of downloading records in ProCite-specific export format thus eliminating the need for a transferring package such as BiblioLinks. Similarly, more powerful hardware has become affordable—extensive searches can be performed much more rapidly. That is not to say, however, that the compilation of a specialized database will be effortless as lack of standardization among the systems will always present problems.

What is of more concern is the issue of copyright and how one can use the information retrieved from these remote databases. Without a doubt, the transferring of machine-readable records into in-house databases will be a trend of the 1990s and beyond. Increasing access to worldwide library collections, full text journals, and bibliographic databases through systems such as Internet, coupled with the advances in the storage of electronic information such as CD-ROM, will invite the creation and maintenance of specialized databases. The thought that these databases can only then be used internally seems inconceivable, especially knowing the amount of time and effort it can take to create such a beast. The issue of "who owns the data?" will have to be addressed sooner or later.

References

Abrami, P.C., & d'Apollonia, S. (1991). Multidimensional students' evaluations of teaching effectiveness—The generalizability of "N=1" research: Comment on Marsh. *Journal of Educational Psychology, 83,* 411-415.

Abrami, P.C., Chambers, B., Poulsen, C., Spence, J. & Lou, Y. (in progress). The effects of group composition on achievement: A meta analysis.

Abrami, P.C., Chambers, B., Poulsen, C., Howden, J., Kastelorizios, K., Wagner, D., & Glashan, A. (1993). *Using cooperative learning.* Dubuque, IA: W.C. Brown-Benchmark Publications.

Abrami, P.C, Chambers, B., Poulsen, C., Kouros, C., Farrell, M., & d'Apollonia, S. (under review). Cooperative learning and social support: A primary investigation and research integration. *Journal of Social Psychology.*

Baron, L.J., & Abrami, P.C. (1992). The effects of group size and exposure time on microcomputer learning. *Computers in Human Behavior, 8,* 353-365.

Duggan, M.K. (1991). Copyright of electronic information: Issues and questions. *Online, 15,* 20-26.

Howe, H., Moller, L., Chambers, B., & Petralos, H. (in press). The ecology of dramatic play centres on children's social and cognitive play. *Early Childhood Education Research Quarterly.*

Schmid, R.F., Abrami, P.C., d'Apollonia, S., DeSimone, C., & Poulsen, C. (in progress). Mental modeling theory applied to the cognitive analysis of learning strategies and tactics.

Williams, M. (1986). Transparent information systems through gateways, front-ends, intermediaries and interfaces. *Journal of the American Society for Information Science, 37,* 201-214.

WilsonDisc, and various other online services. PBS software is also available at various university bookstores at discounted prices.

The author would like to acknowledge the sound advice provided throughout this project by Dr. Philip Abrami, Jacqueline Dealy, and Catherine Poulsen.
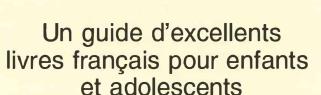
Anne Wade is Manager of the Resource Centre at the Centre for the Study of Classroom Processes (CSCP), Concordia University and also a part-time Reference Librarian at Westmount Public Library, Montreal, Quebec, Canada.

Anne Wade, CSCP, LB-581, 1455 de Maisonneuve Blvd.W., Montreal, Que. H3H 1Y5, Tel: (514) 848-2020, Fax: (514) 848-4520, E-Mail: Wada@vax2.concordia.ca