8-25-1999

Education Policy Analysis Archives 07/25

Arizona State University

University of South Florida

Follow this and additional works at: http://scholarcommons.usf.edu/coedu_pub

Part of the Education Commons

Scholar Commons Citation
http://scholarcommons.usf.edu/coedu_pub/251
The Quality of Researchers’ Searches of the ERIC Database

Scott Hertzberg
ERIC Clearinghouse on Assessment and Evaluation
University of Maryland

Lawrence Rudner
ERIC Clearinghouse on Assessment and Evaluation
University of Maryland

Abstract
During the last ten years, end-users of electronic databases have become progressively less dependent on librarians and other intermediaries. This is certainly the case with the Educational Resources Information Center (ERIC) Database, a resource once accessed by passing a paper query form to a librarian and now increasingly searched directly by end-users. This article empirically examines the search strategies currently being used by researchers and other groups. College professors and educational researchers appear to be doing a better job searching the database than other ERIC patrons. However, the study suggests that most end-users should be using much better search strategies.

A critical component of conducting almost any kind of research is to examine the literature for both related content and previously employed research methods. By reviewing the related literature, researchers are better able to formulate their research questions, build on past research, and design more effective studies. In the field of
education, a usual first step in identifying related literature is to search the over 950,000 citations included in the Educational Resources Information Center (ERIC) database.

With its availability on the Internet and on CD-ROM, the ERIC database is now accessed by a wide and diverse audience and less specialized audience. In May 1999, the ERIC Clearinghouse on Assessment and Evaluation alone had over 3,500 users searching the ERIC database daily. This is quite a change from 10 years ago when access to the ERIC database was typically restricted to trained reference librarians who had accounts with commercial information service organizations such as Dialog.

The question studied in this paper is the quality of the search strategies of today’s end-users. We present effective strategies for searching the ERIC database, a brief summary of the literature on end-user searching, and empirical information on the quality of end-users searches of the ERIC database installed at the ERIC Clearinghouse on Assessment and Evaluation web site.

Effective Strategies for Searching the ERIC Database

The Educational Resources Information Center is the largest source of educational information in the world. The most well-known and frequently used body of information produced by the ERIC system is the ERIC Database which contains close to one million citations and abstracts reflecting both published and "gray literature" (conference papers, contractor reports, etc.) gathered by the 16 ERIC subject area clearinghouses. For over thirty years, the database has been a widely-used and well-known research tool.

The ERIC database can be accessed through various media. Researchers may search the database via Dialog, the Internet or CD-ROMs produced by several vendors. Although, the database is still searchable by way of paper indexes, electronic formats are the concern here because they are largely responsible for the surge in end-user searching.

There are some good search practices that are applicable to all electronic versions of the database. One of the most important tactics is the use of Boolean operators (AND, OR, NOT) to refine queries. One-word and one-phrase searches are rarely sufficient. When using Boolean operators, avoid the common mistake of confusing the function of "AND" and "OR". The query Portfolios AND Nongraded Evaluation retrieves only documents containing both descriptors, while a search for Portfolios OR Nongraded Evaluation retrieves a set of documents that have either or both of the descriptors.

Another fundamental rule for successful searching is to use all relevant descriptors (ERIC indexing terms). Find all related and narrower terms that apply and link them into the search with the Boolean operator "OR". Using all relevant descriptors increases recall (i.e. comprehensiveness of retrieval) and often reveals useful citations not found when searching using only one or two descriptors. The ERIC database is a very well indexed database, but has not been constructed with perfect consistency over the past 30 years. Further, the terms preferred by any individual end-user may not be the same as the terms preferred by the ERIC indexers. For example, ERIC uses Test Wiseness, Student Evaluation and Disadvantaged Youth. The terms Test Preparation, Student Assessment and Disadvantaged Students are not ERIC descriptors. Failing to use the controlled vocabulary terms will result in a search that misses highly relevant documents.

Because of these gaps between the database’s controlled vocabulary and natural language, use of The Thesaurus of ERIC Descriptors (Houston, 1995) is essential to successful searching. The thesaurus, which has been published in paper since the creation of the database, is now available on many CD-ROM versions of the database and uniquely at the website of the ERIC Clearinghouse on Assessment and Evaluation (ERIC/AE).
The thesaurus is incorporated in the Search ERIC Wizard, one of the user interfaces for the ERIC/AE’s Internet version of the database (http://ericae.net/scripts/ewiz/amain2.asp). The ERIC Wizard interacts with users to indicate whether a search term is an actual ERIC descriptor. If a term entered by a user is not a descriptor, the Wizard suggests alternatives. When the correct descriptor is located, the Wizard displays an array of related and narrower terms. The user may then choose from the first term or the related terms to construct a search of the database.

**Hints for Effective Searching**
- Use Boolean operators (AND, OR, NOT) to craft good queries.
- Expand the query by ORing appropriate narrower and related terms.
- Use the print of an electronic ERIC thesaurus to find useful descriptors.
- Use the Building Block or Pearl Building methods.
- Conduct multiple searches.

An added feature of the search engine installed on the ERIC/AE website is a Find Similar link. The Find Similar feature performs a popular search strategy known as Pearl Building. Pearl Building involves the constructing of new searches around descriptors found in the good results of preliminary searches. The Find Similar link for a particular citation will produce a new set of documents that are based on the first document’s descriptors. This function often retrieves useful documents not found in the first search. You can choose the best documents from the second set of citations and continue to re-circulate the search until you no longer find any new, relevant hits. You may also edit the descriptors of a selected document to search only for the descriptors judged relevant to your needs.

Another good technique for organizing a complex search, applicable to all search situations, is the Building Blocks method. On a piece of paper, write out the two or three most essential components of a given question. These are the building blocks of the search. Construct a search by linking the building blocks with what you believe are the correct Boolean operators. If the resultant search is not very successful, expand it by attaching related descriptors to one or more of the building blocks. Continue to add to the building blocks and, if necessary, rearranging the Boolean operators, until you achieve satisfactory results. Inherent in this method is the necessity of conducting multiple queries for a given search.

**Literature Review**

This section summarizes some of the literature with regard to end-user searching with particular attention to the quality of end-user results, quality of search strategies, time spent on a search, use of thesauri, the frequency of multiple searches, and experience. Since this study is concerned with end-user searching of an electronic database through an Internet interface, both studies of users of on-line databases and studies of users of Internet search engines are relevant. Studies of the first type of users are quite numerous, as on-line databases have been widely used for over 20 years. Relevant literature on the search behavior of Internet users, on the other hand, is still rather scarce.

**Quality of end-user results**

There is a large body of literature claiming that most end-users obtain poor results when searching for themselves (Lancaster, Elzy, Zeter, Metzler and Yuen, 1994; Bates
Lancaster, Elzy, Zeter, Metzler and Yuen, for example, compared faculty and student searches of ERIC on CD-ROM to searches conducted by librarians. They noted that most of the end-users found only a third of the relevant articles found by the librarians.

There are several studies, however, where end-users are able to search online databases with good results. Sullivan, Borgman and Wippern (1990) compared the searching of 40 doctoral students given minimal training with searches done by 20 librarians. The 40 students were no less satisfied with their searches of ERIC and Inspec than with the results retrieved by the librarians, and, in fact, found their searches to be more precise. Similarly, the patent attorneys in Vollaro and Hawkin (1986) felt that intermediaries could have done a better job, but were largely satisfied with their own searches. Both studies observed that the end-users still had trouble searching databases. Sullivan, Borgman and Wippern noted that the end-users "made more errors, prepared less well than intermediaries and had less complete results."

There are a few explanations for why some end-users may search more successfully than others. Yang (1997) observed that certain concepts and metaphors used by novice users to construct searches were beneficial to searching. Marchionini, Dwiggins and Katz (1993) suggested that subject expertise helps end-users search more effectively.

**Strategies**

Several studies have concluded that end-users use poor searching techniques, marked by overly simple statements and limited use of Boolean operators or other commands (Bates and Siegfried, 1993; Tolle and Hah, 1985; Teitelbaum and Sewell, 1986). In their study of 27 humanities scholars, Bates and Siegfried (1993) observed that 63% of the searches contained only one or two terms and 25% included no Boolean operators at all.

Nims and Rich (1998) studied over 1,000 searches conducted on the Search Voyeur webpage hosted by Magellan. The Search Voyeur site allows users to spy on the searches of other users. The researchers found a profusion of poorly constructed searches. Searchers performed one-word searches when more complex queries linked with Boolean operators were necessary. Overall, a mere 13% of the searches used Boolean operators. The study, which observed how the general public searches the entire World Wide Web, suggests that end-users may have more trouble searching Internet databases than older online databases. End-users of Internet databases may be less familiar with the search protocols and may have higher expectations of the technology’s ability to make up for their poor searching techniques.

**Time Spent Searching**

Looking at the transaction logs of 11,067 search sessions on computers linked to Medline at the National Library of Medicine, Tolle and Hah (1985) found that end-users averaged significantly less time searching than librarians. Patrons in the study averaged 15 minutes of searching per session, while librarians in the control group averaged 20 to 25 minutes.

**Use of a Thesaurus**

In their study of 41 patent attorneys searching Inspec, Valloro and Hawkins (1986) observed that the majority of the end-users did not utilize the database’s thesaurus. Interviews revealed that most of the subjects did not feel familiar enough with the main functions of the database to effectively use the thesaurus (which they considered an
advanced feature). The study suggests that end-users may be under-utilizing online thesauri, but the subject remains largely unexamined.

**Number of Queries**

Conducting multiple searches is often essential to successful searching. Yet studies suggest that only around half of all end-users perform more than one search per session. (Spink 1996; Huang 1992). Spink conducted 100 interviews with academic end-users at Rutgers University and found that only 44% conducted multiple searches per session.

**Experience**

The most significant factor determining searching success appears to be experience using a database. In a recent study of law school students searching Quicklaw, Yuan (1997) showed that the search repertoires of students became more complex and effective over time. Tolle and Hah (1986) found a correlation between experience and the frequency of multiple searches. Only 8% of the experienced users in the study stopped searching after a failed search, while the rate of stopping was 11% for moderately experienced users and 20% for inexperienced users.

**Summary**

The quality of end-user searching appears to vary depending on the individual end-user. Some searchers are stronger than others because of skills they bring to searching or gain from using an online database over time. However, the literature suggests that most end-users could be doing better. Even the studies that recorded a high level of end-user satisfaction, observed that end-users rely on overly simple searches, make frequent errors, and fail to attain comprehensive results.

**Method**

For two days in early November 1998, all patrons wanting to search the ERIC database installed at the ERIC/AE website were required to complete a 10-item background questionnaire. For each patron, we then tracked a) the maximum number of OR’s in their searches as a measure of search quality, b) the number of queries per session, c) whether they used the thesaurus or free-text search engine, d) number of hits examined, and e) the amount of time devoted to searching the ERIC database per session.

Data were collected on 4,086 user sessions. Because some browsers were not set to accept identifiers, we were not always able to relate background data to session information. Accordingly, our analysis is based on the 3,420 users with background and corresponding session information.

Participation in the study was entirely voluntary; patrons could go elsewhere to search the ERIC database. However, our questionnaire was short and our data collection was unobtrusive. Based on the prior week’s log, we estimate our retention rate was over 90%.

**Results**

We asked our end-users "what is the primary purpose of your search today?". As shown in Table 1, most patrons were searching in connection with preparing a research report.

| Table 1 |
Purpose of searching the ERIC database

<table>
<thead>
<tr>
<th>Purpose</th>
<th>N</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research report preparation</td>
<td>1825</td>
<td>53.4%</td>
</tr>
<tr>
<td>Class assignment</td>
<td>601</td>
<td>17.6</td>
</tr>
<tr>
<td>Professional interest</td>
<td>554</td>
<td>16.2</td>
</tr>
<tr>
<td>Lesson planning</td>
<td>177</td>
<td>5.2</td>
</tr>
<tr>
<td>Background for policy making</td>
<td>175</td>
<td>5.1</td>
</tr>
<tr>
<td>Classroom management</td>
<td>88</td>
<td>2.6</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>3240</strong></td>
<td><strong>100.0%</strong></td>
</tr>
</tbody>
</table>

Some searching characteristics of the entire sample and of groups of individuals who identified themselves as college librarians, college professors, and researchers are presented in Table 2. College librarians are presumably the most trained and most experienced user group, while college professors and researchers are presumably the most diligent user group.

Most variables were fairly normally distributed. Accordingly, means and standard deviations (std dev) are presented in the table. The amount of time spent searching, however, was quite skewed. Central tendency and variability for time are represented by medians and semi-interquartile ranges (sir).

Table 2
Searching Characteristics for Select User Groups

<table>
<thead>
<tr>
<th>Quality</th>
<th>N queries</th>
<th>Thesaurus Use</th>
<th>Hits Examined</th>
<th>Time (in seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>Mean</td>
<td>Std dev</td>
<td>Mean</td>
<td>Std dev</td>
</tr>
</tbody>
</table>

College Librarian
- 96
- .91
- 3.89
- 2.66
- 3.26
- 46.8
- 3.11
- 5.41
- 207
- 240

Researcher
- 445
- .42
- 1.26
- 3.04
- 3.69
- 37.6
- 4.85
- 10.23
- 376
- 408

College Professor
- 209
- .37
- 1.10
- 2.49
- 2.46
- 44.6
- 5.58
- 15.09
- 361
- 345

All users
- 3420
- .44
- 1.77
- 2.75
- 2.95
- 38.7
- 3.65
- 8.65
- 352
- 351

A good search incorporates Boolean operators to capture appropriate terms. As a
measure of search quality, we noted the maximum number of OR’s used in any query during a patron’s session. The data indicate that there is about one OR in every two search sessions. College librarians tend to conduct the most complicated searches and college professors conducted the simplest searches. To provide an additional perspective on these numbers, we computed the number of OR’s used in the 84 pre-packaged search strategies at http://ericac.net/scripts/ewiz/expert.htm. These search strategies were developed by the top reference librarians across the entire ERIC system. The mean number of OR’s used in these high quality, general purpose searches was 2.9 with a standard deviation of 2.8. Thus, the data show that on-line users tend to be conducting very simple searches that do not take account of subject matter nuances.

The typical user performs 2 to 3 queries per search session and there is little variability across groups. In contrast, the reference staff at the ERIC Clearinghouse on Assessment and Evaluation typically conduct 3 to 6 searches when responding to patron inquiries.

Not using the ERIC thesaurus to guide a search is equivalent to guessing which terms are used by the ERIC indexers. Using the thesaurus, one can employ the proper terms in a search. College librarians and college professors use the thesaurus much more often than most users. Yet, less than half of the searches at the ERIC/AE site take advantage of this unique, special feature.

For any given topic in education, there is typically a large number of related papers and resources. To find all the resources which meet their specific purposes, users need to examine a large number of citations. College professors and researchers are much more diligent than other users in examining citations. Further, as noted by the variance, some professors and researchers are looking at a very large number of citations. Still, the average number of citations examined is quite small, typically about 5 or 6 hits for the most diligent groups. It appears that most patrons, especially those that are not trained researchers, are not looking beyond the first page of hits.

The study showed that the median amount of time spent searching the ERIC/AE site is about 6 minutes. College professors and researchers spend slightly more time than the typical user searching for information. College librarians spend considerably less time searching.

At a minimum, we would like to see at least one OR in the query, more than one query, and at least four hits examined. Only 153 (4.5%) of our examined 3420 users met these criteria.

Discussion

Our findings with regard to Internet searching of the ERIC database are consistent with the broader literature on end-user database searching. Some researchers may be doing a better job than most patrons. Nevertheless, most end-users are conducting few searches, crafting poor searches, not using the thesaurus, and are examining only a few potential hits. While there are times an end-user may want to quickly look up something, such as finding a reference, research report preparation usually involves finding a collection of several relevant, high quality studies. This work cannot be done quickly. Ninety-five percent of the searches we examined do not meet our minimal criteria. From our point of view, these results are very disappointing. Patrons are not using effective search strategies and cannot possibly find the best and most relevant articles in the database being searched.

We have reason to believe that most end-users are satisfied with any somewhat-relevant hit and are not looking for the best citations. After we added the Find
Similar option to our search engine, we noted that few end-users were taking advantage of the feature. We posted a short survey for a few hours asking why. The vast majority of users (80%) told us they were able to find what they wanted on the first page of hits. The reality is that with the default search options, hits are presented in what is basically chronological order. The ranked relevance option does not necessarily present the best quality documents first. Users may be satisfied, but they are not finding the best.

We cannot place enough emphasis on the need to use the Thesaurus of ERIC Descriptors when constructing a search strategy. In addition to the need to include related and narrower terms, the philosophy behind the ERIC Thesaurus and its structure necessitate added diligence on the part of the searcher. The ERIC Thesaurus is designed to reflect the terms used in the professional and scholarly education literature. It is not a strictly hierarchical thesaurus with a rigid set of mutually-exclusive term arrays. Thus, the ERIC Thesaurus is populated with terms that partially overlap and its structure sometimes necessitates variable search strategy design. For example, to find the documents that address the evaluation of instructional methods or activities one should search "Course Evaluation" OR "Curriculum Evaluation". This is a problem with the social sciences in general as terms are less well defined, more fluid and less strictly hierarchical than in the physical sciences.

We occasionally hear frustration from the research community with regard to the ERIC database. The data imply that much of the end-user frustration is due to poor end-user searches. This is not to say the ERIC database is not without its faults. The ERIC system has basically been level-funded for the past 20 years and there has been no system-wide examination of ERIC’s acquisition and processing efforts in 20 years. As a result, there are gaps in ERIC coverage. At our own clearinghouse, we have noted that the 39 journals that we process for inclusion in the ERIC database produce 1,100 articles. Yet, due to our budget, we have usually been limited to entering 700 articles per year. We process few international journals and are slow to add new journals, regardless of their quality or prominence.

We believe there has also been a steady decline in the "gray" literature portion of the ERIC database. Of the approximately 5,500 papers presented at the annual meetings of the American Educational Research Association, for example, only about 1,200 are entered into the ERIC database. Many authors do not have prepared papers and many that have papers do not respond to solicitation requests. Authors should view ERIC as a reproduction service. We make copies of papers available to others. Inclusion in the ERIC database only means that a paper has met some minimal acceptability criteria; it is not equivalent to peer-reviewed publishing and it should not preclude an author from submitting their paper to a refereed journal. Accordingly, we do not see any reason an author should not submit their paper to ERIC. In fact, submitting high quality papers can result in more people seeing the research and more people submitting their papers. Thus, we believe many authors are not assuming their share of the responsibility in building the ERIC resource.

While ERIC database content has its limitations, we believe the lack of end-user search skills is the major impediment to locating the best and most relevant resources. Poorly formed searches and poor search strategies cannot possibly find the best citations. We are encouraged by the conclusions of Sullivan, Borgman and Wippern (1990). With minimal training and a bit of diligence, end-users can attain satisfactory results. It is our hope that readers of this article will follow the suggestions outlined at the beginning of this paper and, concomitantly, increase their chances of finding the best and most relevant documents in the ERIC database.
Note

We wish to thank Dagobert Soergel, Jim Houston and Ted Brandhorst for their useful suggestions on an earlier version of this paper.

References


About the Authors

Scott Hertzberg is a Research Assistant at the ERIC Clearinghouse on Assessment and Evaluation, College of Library and Information Services, 1129 Shriver Laboratory, University of Maryland, College Park, Maryland, 20742. He specializes in social science information services.

Lawrence Rudner is the Director of ERIC Clearinghouse on Assessment and Evaluation, College of Library and Information Services, 1129 Shriver Laboratory, University of Maryland, College Park, Maryland, 20742. He specializes in assessment and information services. He can be reached at rudner@ERICe.net.

Copyright 1999 by the Education Policy Analysis Archives

The World Wide Web address for the Education Policy Analysis Archives is http://epaa.asu.edu

General questions about appropriateness of topics or particular articles may be addressed to the Editor, Gene V Glass, glass@asu.edu or reach him at College of Education, Arizona State University, Tempe, AZ 85287-0211. (602-965-9644). The Book Review Editor is Walter E. Shepherd: shepherd@asu.edu. The Commentary Editor is Casey D. Cobb: casey.cobb@unh.edu.

EPAA Editorial Board

Michael W. Apple
University of Wisconsin

John Covaleskie
Northern Michigan University

Alan Davis
University of Colorado, Denver

Mark E. Fetler
California Commission on Teacher Credentialing

Thomas F. Green
Syracuse University

Arlen Gullickson
Western Michigan University

Aimee Howley
Ohio University

William Hunter
University of Calgary

Daniel Kallós
Umeå University

Greg Camilli
Rutgers University

Andrew Coulson
a_coulson@msn.com

Sherman Dorn
University of South Florida

Richard Garlikov
hmwkhelp@scott.net

Alison I. Griffith
York University

Ernest R. House
University of Colorado

Craig B. Howley
Appalachia Educational Laboratory

Richard M. Jaeger
University of North Carolina—Greensboro

Benjamin Levin
University of Manitoba
Thomas Mauhs- Pugh  
Green Mountain College

Dewayne Matthews  
Western Interstate Commission for Higher Education

William McInerney  
Purdue University

Mary McKeown-Moak  
MGT of America (Austin, TX)

Les McLean  
University of Toronto

Susan Bobbitt Nolen  
University of Washington

Anne L. Pemberton  
apembert@pen.k12.va.us

Hugh G. Petrie  
SUNY Buffalo

Richard C. Richardson  
New York University

Anthony G. Rud Jr.  
Purdue University

Dennis Sayers  
Ann Leavenworth Center for Accelerated Learning

Jay D. Scribner  
University of Texas at Austin

Michael Scriven  
sriven@aol.com

Robert E. Stake  
University of Illinois—UC

Robert Stonehill  
U.S. Department of Education

Robert T. Stout  
Arizona State University

David D. Williams  
Brigham Young University

EPAA Spanish Language Editorial Board

Associate Editor for Spanish Language
Roberto Rodríguez Gómez  
Universidad Nacional Autónoma de México

roberto@servidor.unam.mx

Adrián Acosta (México)  
Universidad de Guadalajara  
adrianacosta@compuserve.com

J. Félix Angulo Rasco (Spain)  
Universidad de Cádiz  
felix.angulo@uca.es

Teresa Bracho (México)  
Centro de Investigación y Docencia Económica-CIDÉ  
bracho dis1.cide.mx

Alejandro Canales (México)  
Universidad Nacional Autónoma de México  
canalesa@servidor.unam.mx

Ursula Casanova (U.S.A.)  
Arizona State University  
casanova@asu.edu

José Contreras Domingo  
Universitat de Barcelona  
Jose.Contreras@doc.d5.ub.es

Erwin Epstein (U.S.A.)  
Loyola University of Chicago  
eepstein@luc.edu

Josué González (U.S.A.)  
Arizona State University  
josue@asu.edu

Rollin Kent (México)  
Departamento de Investigación Educativa- DIE/CINVESTAV  
rkent@gemtel.com.mx  
kentr@data.net.mx

María Beatriz Luce (Brazil)  
Universidad Federal de Rio Grande do Sul- UFRGS  
lucemb@orion.ufrgs.br