# The Search Value Added by Professional Indexing to a Bibliographic Database†

Philip Hider

Charles Sturt University, School of Information Studies,
Boorooma Street Wagga Wagga, New South Wales 2678 Australia,
<phider@csu.edu.au>

Professor Philip Hider has been Head of the School of Information Studies at Charles Sturt University (CSU), Australia since 2008. Before arriving at CSU in 2003, he headed up the Singapore bibliographic network, and, before that, worked as a cataloguer at the British Library. Professor Hider holds a PhD from City University, London and served on the Australian Committee on Cataloguing from 2004-2009. His scholarly interests centre around knowledge and information organisation, having published, taught, researched and practised in the field for well over two decades.

Hider, Philip. 2017. "The Search Value Added by Professional Indexing to a Bibliographic Database." *Knowledge Organization* 44(8): 23-32. 20 references. DOI:10.5771/0943-7444-2018-1-23.

**Abstract:** Gross et al. (2015) have demonstrated that about a quarter of hits would typically be lost to keyword searchers if contemporary academic library catalogs dropped their controlled subject headings. This article reports on an investigation of the search value that subject descriptors and identifiers assigned by professional indexers add to a bibliographic database, namely the *Australian Education Index* (AEI). First, a similar methodology to that developed by Gross et al. (2015) was applied, with keyword searches representing a range of educational topics run on the AEI database with and without its subject indexing. The results indicated that AEI users would also lose, on average, about a quarter of hits per query. Second, an alternative research design was applied in which an experienced literature searcher was asked to find resources on a set of educational topics on an AEI database stripped of its subject indexing and then asked to search for additional resources on the same topics after the subject indexing had been reinserted. In this study, the proportion of additional resources that would have been lost had it not been for the subject indexing was again found to be about a quarter of the total resources found for each topic, on average.

Received: 29 August 2017; Revised: 16 October 2017; Accepted 30 October 30 2017

Keywords:

† The author wishes to thank the ACER Cunningham Library for facilitating this project and Tarita Dickson for her assistance with the second study.

## 1.0 Introduction

While indexers and catalogers might complain that empirical evidence pointing to the value their work adds to databases and catalogs is not always noted, or given much weight, by their employers, it is important that this evidence continues to be collected and reported, just as it is important that any evidence that suggests a decline in the value of professional indexing and cataloging is likewise reported. Metadata and KO professionals may need to consider a range of survival strategies in a "post-truth world," including those suggested by Gross (2015) and Borie et al. (2015), but they first need to confirm, if only to themselves, the continued value of their work; this is best done through a thorough, and open, engagement with the data.

The research described in this paper follows up on the studies conducted by Gross and Taylor (2005) and Gross et al. (2015), which provided evidence for the ongoing value of subject headings in a contemporary academic library catalog, i.e., that of the University of Pittsburgh. They found that, on average, about a quarter of "hits" in real-life keyword searches would not have been retrieved were it not for one or more subject headings, even after the catalog had been enriched with tables of contents and other derived indexing. The subject headings, of course, would been assigned, in almost all cases, by professional catalogers. The findings suggest that this key component of catalogers' work, i.e., subject indexing, continues to significantly assist library patrons in their subject searching, at least in relation to their use of catalog data (even if not in the catalog itself), and their need for the resources that data represents.

Although the two studies by Gross et al. make a number of assumptions, as will be discussed shortly, they are

based on a relatively straightforward methodology that can be readily replicated, and this article first reports on the findings of a similar study that examined the impact on retrieval of another branch of professional indexing, namely, that carried out for periodical and bibliographic databases. Again, it focuses on the value specifically of assigned subject indexing. A second study with similar objectives is then reported, which was based on a before-and-after experimental research design that aimed to address some of the limitations of the methodology employed by Gross et al.

While subject indexing is generally regarded as one of the most important and "professional" activities performed in cataloging, it is typically an even more central activity in database indexing. If it was found to add little value, then the case for the professional database indexer would surely be weak. Conversely, if database searches are much assisted by professionally assigned subject indexing that could not be readily assigned by authors or other non-professionals, then the case for professional intervention would be intrinsically strong.

## 2.0 Literature Review

The value of assigned indexing, and in particular assigned indexing using controlled language, was first called into question with the publication of results from the "Cranfield" experiments, which found that, for topical document retrieval, certain forms of derived indexing could achieve higher recall and precision ratios than those achieved by the various controlled vocabularies tested (Cleverdon 1967). Numerous studies and discussions of the relative merits of controlled and derived indexing since have pointed to the "received wisdom" of the two approaches' complementarity, each with strengths and weaknesses more or less exposed in different retrieval contexts (Rowley 1994; Bawden and Robinson 2012). The question remains, however, as to whether the value that controlled indexing, particularly of the sophisticated kind undertaken by information professionals, adds to a given search context is sufficiently large to justify its costs. This has recently been addressed by Gross and Taylor (2005) and Gross et al. (2015) in the context of the academic library catalog. The reality of this environment is not yet one of comprehensive "full-text retrieval" (that is, retrieval based on full-text indexing), but rather of retrieval (for resources on topics) based, predominantly, on titles, tables of contents, summaries and limited amounts of other "content," along with cataloger-assigned subject headings. Gross et al. (2015) found that the number of records retrieved in the University of Pittsburgh's library catalog by keyword (subject) searches, that were "only" retrieved because of the inclusion of one or more subject headings, represented, on average, about a quarter of the total records retrieved. Such a proportion might be considered insufficiently large to warrant the expense of professional subject indexing in the case of "casual" searching, but proponents of detailed cataloging argue that "scholarly" searching requires more comprehensive results (Gross et al. 2015; Mann 2008).

While many experiments have been carried out to evaluate the effect of controlled subject vocabularies in bibliographic databases (indeed, more than in library catalogs; some of the earlier key studies are cited in Rowley's 1994 survey; more recent examples include those reported by Kim (2014), Savoy (2005) and Muddamalle (1998)), the methodology employed by Gross et al. (2015) does not appear to have been replicated in this particular environment. Bibliographic databases are defined here as the products of the various journal indexing services, which sometimes also support direct (online) access to full texts. It is unclear whether the subject indexing provided by these services, often based on a thesaurus, enhances retrieval to a similar extent, to that of, say, *Library of Congress Subject Headings* (*LCSH*) in library catalogs. There are a number of differences between the two kinds of environment that might affect the indexing's relative impact, as indeed there are across individual bibliographic databases and individual catalogs, such as the nature and quantity of other data elements, including abstracts, present in the keyword index, and the breadth and depth of the controlled vocabulary (if used) relative to the breadth and depth of the topics searched for by the database or catalog users.

The methodology of Gross et al. has some limitations, however. As they themselves point out, it allows for a measure of "hits" lost, but is silent on whether or not these hits are "relevant" (Gross et al. 2015). They speculate that the proportion of "hits" that are not relevant is likely to be less on catalogs with subject headings than catalogs without them, as precision tends to be a strength of controlled vocabularies, although this has yet to be demonstrated.

Moreover, the measure provided by the methodology does not necessarily reflect "actual" retrieval loss, because it is based on "individual" search results, i.e., from a single query, whereas in real life users may perform "follow-up" searches, based on their pre-existing knowledge and/or on ideas for search terms triggered by their initial interaction with the search system, e.g., in displayed records. In relation to the findings of Gross et al., on the one hand searchers may reduce loss levels by following up using *LCSH* terms independently of any *LCSH* they encounter; on the other, searchers may increase loss levels by following up using *LCSH* terms that they derive from the *LCSH* they encounter, as, in that case, a relatively high proportion of additional hits would not have

occurred without the *LCSH*. It is unclear which of these opposing phenomena might be larger.

There can be no doubt, however, that iterative searching takes place and is a significant factor in document retrieval (Hider 2006; Rieh and Xie 2006; Zhang 2013; Zhang and Soergel 2014; Pontis and Blandford 2015). Hjørland (2014, 1563) notes how search iteration and "selection power" may be particularly important for advanced users carrying out systematic searches, who approach their searching "hermeneutically" in which "there is a constant reinterpretation of the relevant literature" and a need to understand "what is going on during the search." Hjørland (2014) goes on to argue that "classical databases," with an underlying Boolean retrieval model, are especially suited to the hermeneutical approach, which he contrasts to the "positivist" approach, on which the alternative statistical models, developed later by the (automated) "information retrieval" field, are based. The database studied in the research reported here is, in this sense, an example of a "classical database."

While some search iteration may be due to searchers being aware at the start of their searching of synonyms and the value of trying multiple search strategies, they themselves would not always be able to fully address the so-called "synonym problem" (Weber et al. 2006). Indeed, one wonders how often searchers can think of "all" possible word forms of "all" synonyms and near-synonyms, in "all" languages. Subject headings and thesauri not only limit this problem but also suggest search terms for related concepts that might well unearth other relevant resources. This can happen either "pre-hoc," e.g., through preliminary thesaurus consultation, or "post-hoc," e.g., through links in records and subject facet displays. As Hider (2017) has recently pointed out, professional cataloging can add value across a wide range of catalog user tasks and not just to the tasks that involve retrieving bibliographic records via a generic search box. These other user tasks include those associated with the consultation of subject authority files and those associated with "browsing."

The studies by Gross et al. (2015) also raise questions, as mentioned earlier, around the "interpretation" of the measure of search value they employed. In particular, at what level does retrieval loss become "bad?" In some search contexts, there may be little need for a high recall ratio, where relatively few, reasonably relevant resources may suffice; whereas in other search contexts, the objective might be full recall, or the user may be significantly disadvantaged if, say, one out of four relevant resources was missed. The "need" for resources, and particular recall levels, are themselves very difficult things to measure; indeed, they have yet to be convincingly measured, which is one of the reasons why there is no definitive answer to the relative values of controlled and derived indexing.

## 3.0 Design of First Study

The first study of the reported research project applies the methodology developed by Gross et al. (2015) to a particular bibliographic database, namely, the *Australian Education Index* (AEI), which "provides a complex and sophisticated subscription database consisting of more than 200 000 entries relating to educational research, policy and practice" (ACER Cunningham Library 2017a). The database covers predominantly English-language material. The professional indexers who maintain AEI assign subject terms from the *Australian Thesaurus of Education Descriptors* (ATED http://cunningham.acer.edu.au/multites2007/index.html), along with identifiers and geographic names where applicable. ATED includes "over 5,000 main entry descriptors," along with many cross-references, and "reflects terminology used to describe research and practice in Australian education" (ACER Cunningham Library 2017b). AEI records also include the titles and subtitles, abstracts and journal names of the articles indexed, all of which may provide an indication of subject. However, they do not include author-assigned "keywords" (although such terms are sometimes used by the indexers to assist their subject analysis).

Whereas in the studies by Gross et al., the proportion of resources that would be lost without the subject indexing was estimated by analyzing, in some cases manually, the content of the records retrieved from searches on the library catalog, it was possible to calculate the "loss levels" for this study simply by running the sample of search queries twice: first on all the "basic keyword" indexes, and then on all the "basic keyword" indexes "except" for those containing the assigned subject terms.

The sample of queries used in the study was derived in a similar, though not identical, way to that of the studies by Gross et al. (2015) in which a set of search terms was derived from the catalog system's transaction log. In the Gross studies, after duplicate terms were removed, every (presumably chronologically) tenth term was taken for the sample, except if it had resulted in no hits or more than 10,000 hits. The AEI study did not have access to any search logs from the AEI database itself but was provided with a recent transaction log of (general) keyword searches on EdResearch Online (http://opac.acer.edu.au/edresearch), which is based on AEI and provides access to "over 56 000 articles from more than 500 Australian education journals" (ACER Cunningham Library 2017c). To maximise the sample size but without also ending up with a large proportion of interdependent queries (that is, queries from the same series of searches on a topic) every (chronologically) fifth search query was taken from the de-duplicated log. The resulting set of queries was found, however, to include a large number of

queries that were clearly not "topics," representing instead searches for known articles, journals, authors, etc. These were duly identified and eliminated.

While Gross et al. (2015) had excluded those terms resulting in more than 10,000 hits for practical reasons, the author decided that there were also theoretical grounds for excluding overly vast result sets from analysis: it was thought unlikely that researchers and scholars, or their assistants, would typically wade through quite so many records, even for "thorough" literature searches, and, even if they had immediate access to the full texts, would likely stop inspecting a results set at a certain point, and limit the results set to a more manageable size or conduct a different search. The EdResearch Online log recorded the queries' hit numbers, and these were used as a guide to the number of hits one might expect, for a given query, on AEI (i.e., up to about five times as many). It was decided to exclude those queries with more than 100 hits in the log, so that only those queries likely to yield substantially fewer than 1,000 hits on AEI would be included. It was felt that a very thorough research assistant may be prepared to look through entire result sets if they numbered in the hundreds but not in the thousands.

Although queries with zero hits in the EdResearch Online log might have yielded some hits on the AEI database, it was decided to exclude these as well, along with those with more than 100 hits, so that the final sample size numbered sixty-three. This made it considerably smaller than the 191 search terms analysed in the later study by Gross et al. (2015), but it was considered adequate for the purposes of providing indicative results, providing for a confidence level of just over 70% with a 4% margin of error, if the statistical formula used by Gross et al. (2015) is applied. It should be noted that the queries were left in their natural (i.e., original) state, which meant that a few incorporated the Boolean logical operator "AND" or truncation. The sample queries are listed in Appendix A.

## 4.0 Results of First Study

The effect of the omission of the subject indexes on the sixty-three keyword searches is detailed in Appendix A. The percentage of lost hits across the sample ranges from zero to 78.1%, with a mean of 27.0% and a median of 23.3%. Interestingly, the mean "matches" that produced by the later study of Gross et al. (2015) for all-language materials; the corresponding median was 17.6%. Overall, the sample of queries retrieved 5,256 hits with the subject indexes and 3,898 without them, representing a percentage loss of 25.8%. This compares with a loss of 27.7% in the later study by Gross et al. (2015) for all-language materials. Nine of the sixty-three queries lost 50% or more of their hits without the subject indexes, suggesting that half or more hits would be lost from one in every seven "successful" subject searches on the database. This compares with one in every five searches in the University of Pittsburgh catalog (Gross et al. 2015). In summary, the analysis indicates that similar loss levels, with respect to subject searching, might be expected if the AEI database and the University of Pittsburgh library catalog were not supported by professional indexing.

## 5.0 Design of Second Study

Whether users of the AEI database really do miss out on about a quarter of relevant resources when subject searching is not fully answered by the first study, given the methodological limitations outlined earlier. An alternative research design based on a before-and-after experiment was developed to address some of those limitations and employed in a second study.

In the new study, a research assistant, with experience in the field of education as well as in reference librarianship, was provided with a list of topics that an academic might wish to engage an assistant to search for on the AEI, derived (as detailed later) from the list of queries used in the initial study. For the first stage, the assistant was asked to conduct her literature searches using a version of the AEI stripped of its assigned subject terms (as well as its subject search option on the advanced interface), and to find as many relevant, or potentially relevant, articles as possible, with no limit placed on the number of searches she could try (for practical purposes, a time limit of one hour per topic was imposed), and to compile a bibliography for each topic. The standard DBTextWorks search interface used by the ACER Cunningham Library to provide its local clientele with access to AEI was reconfigured accordingly and connected to a remote login disclosed only to the author and research assistant. The assistant could make use of all search functionality available, including links to full text, as she saw fit; she was not advised, at this stage, that the database had been stripped of its subject indexing but was instead advised that the search interface was experimental.

The same research assistant was then asked, for the second stage, to find any "additional" resources that she deemed relevant, or potentially relevant, for each of the topics previously searched for, on another version of the AEI database, this time with the assigned subject indexing, and subject search option, reinserted. She was advised that the database had been enhanced accordingly. The research assistant was asked to re-enter all the basic keyword queries she had performed earlier, according to a list of them derived from the transaction log. She was invited to also enter any additional queries, as she saw fit,

based on any assigned subject indexing she encountered in retrieved records, or based on other new bibliographic information retrieved from those additional searches. Likewise, she was also invited to click on any additional links she encountered, as she saw fit, including (and in particular) those for subject descriptors.

The research assistant could, in the second stage as in the first, make use of all search functionality available, although this did not include any facet displays or thesaurus look-up. She was asked to add entries for any new resources she found to the bibliographies. For practical purposes, the assistant was given another hour per topic.

Twenty topics were derived directly from the sample of real-life queries used in the first study, selected as those thought likely to be clearest (as topics) to the research assistant, based on the criteria of unambiguousness (some terms had plural meanings or shades of meaning) and lacking in jargon (with which the assistant might not be familiar). They are listed in Table 1. Although the sample size was small, it was considered large enough to yield an indicative measure of retrieval loss, given the exhaustive nature of the searching. A premise of this study, as of the first study, is the need for "total" recall: in high-quality scholarship, literature reviews should be based, it is assumed, on an examination of "all" available resources on their topics.

Apart from the small sample size, there are certain other methodological limitations to this study that should be noted. First, although the research assistant was asked to apply the same threshold for "potential relevance" in the second stage as in the initial stage, it is possible that it was lowered as part of a "Hawthorne effect," in which the assistant felt obliged to try harder to find additional resources than she had tried to find initial resources. On the other hand, the repetition of searching on the same topics and queries might have fatigued the assistant, resulting in a decline in interest and effort. An experimental design that would avoid these issues might involve two research assistants tasked to search both "before" and "after" databases, but only on one of them for each topic, such that the two databases were used by different searchers for each topic. However, the number of topics that allowed for an appropriate level of confidence in a reasonably narrow range of percentage retrieval loss, would need to be sufficiently large to address certain biases that the introduction of a second research assistant would entail: in particular, differences in the two assistants' search skills and techniques, relevance judgments, including thresholds, and topic knowledge.

Second, the activities carried out by the research assistant in the first and second stages were not identical, yet the same time limit was employed. By the same token, the activities were similar (and the topics the same), such that the assistant could have speeded up in the second stage, or perhaps, due to fatigue, slowed down. Ideally, the assistant would have completed their "exhaustive" searches, in both stages, before the time limit, circumventing these issues, but that was not always the case here. In some circumstances, it might be possible for the assistant to be allocated the work without any time limit, as part of a larger, fixed contract. The issues are likewise avoided if two assistants are employed in the experimental design outlined above.

Third, the relevance judgements in the study were made solely by the assistant, who possessed knowledge of the field but not expertise across all the various topics. In real life, the client might have given their assistant suggested search strategies, as well further insights into their specific information needs, affecting both her search behavior and relevance judgements. It is possible that this scenario could be accommodated by an experiment in which both scholars, with genuine literature search requests, and assistants are recruited. This would also address the limitation of using search queries as "topics," although it would compound the biases associated with the alternative experimental design, outlined above, involving interchanged searchers.

Finally, it should be recognized that different searchers, even amongst those with relatively advanced search skills and a fair degree of knowledge of the subject field, will not necessarily perform literature searches, not even exhaustive searches, with the same level of success. The reported experiment assumes that the variation of search skills and subject knowledge in this context would not overly affect the outcome, or at least that the assistant employed for the study possessed skills and knowledge that were reasonably typical. If the alternative experimental design outlined above is carried out, it would be worth incorporating an additional test in which the two searchers were asked to perform exhaustive searches for additional, but the same, topics, on the same version of the database, to gauge the level of inter-searcher reliability. Even if a high level of reliability was determined, the experiment would ideally be replicated using multiple searchers (on the same database). Of course, this would require considerable resources that may not be available.

## 6.0 Results of Second Study

Table 1 shows the number of potentially relevant resources found in the initial stage of the study, when the AEI database was stripped of its subject indexing, and the additional resources found with the aid of the reinserted subject indexing. For two topics (10%), there were no additional resources found; in contrast, for two other topics, more than twice the number of initial resources

| *Topic* | *Initial finds* | *Additional finds* | *Increase (%)* |
|---|---|---|---|
| Sexual assault on campus | 1 | 3 | 300.0 |
| Dealing with angry and aggressive children | 71 | 10 | 14.1 |
| Cueing systems & reading | 7 | 0 | 0.0 |
| Animal assisted therapy | 4 | 1 | 25.0 |
| Learning disabilities & brain research | 46 | 5 | 10.9 |
| Importance of science in primary education | 92 | 22 | 23.9 |
| Physical activity and the academic performance of children | 25 | 3 | 12.0 |
| Giftedness in music | 37 | 2 | 5.4 |
| Guided reading for children with learning difficulties | 90 | 7 | 7.8 |
| Characteristics of first generation college students | 26 | 1 | 3.8 |
| Selective school - year 9 | 11 | 0 | 0.0 |
| Sensory play | 17 | 37 | 217.6 |
| Assessment of large online-distance cohorts - higher education | 34 | 19 | 55.9 |
| Flipped learning | 34 | 1 | 2.9 |
| New arrival programs for immigrants | 26 | 7 | 26.9 |
| Saturday school | 18 | 1 | 5.6 |
| Gender segregation in schools | 6 | 1 | 16.7 |
| Art therapy | 26 | 11 | 42.3 |
| Digital storytelling | 65 | 17 | 26.2 |
| Cyberbullying | 106 | 5 | 4.7 |

*Table 1*. Resources found before and after subject indexing.

were found via the subject indexing. Clearly there was a wide variation in the effect of the subject indexing on recall. Overall, the mean increase in resources found, across the twenty topics, was 40.1%, and the median increase 13.0%. This represents a mean percentage "loss," were it not for the subject indexing, of 28.6%, and a median loss of 11.5%. It is interesting that the mean loss, per topic, is very close to that of the first study (and the Gross study) of 27.0%, suggesting that the impact of the AEI subject indexing on single-query and iterative searching may be similar, at least in terms of recall.

The number of words in the twenty topics, as articulated, ranged from one to seven, excluding "stop words" (e.g., "and," "on," "with"), and allowed for an analysis of the correlation between this variable and the impact of the subject indexing on recall. While Gross et al. (2015) found a possible (but not statistically significant) correlation between numbers of keywords in the catalog searches and hits lost without the *LCSH*, the Spearman's rank correlation coefficient for the two variables in this study (albeit with a small sample size) was a very weak, and negative, -0.11.

## 7.0 Discussion

The results of the two studies reported in this article suggest that professional indexing (in the narrower sense) is as important for scholarly subject searching as professional cataloging is. There can be no universal threshold above which professionally assigned subject indexing is economically justifiable, as this is dependent on competing claims to funds and the nature of the subject searching performed on any given database. Indeed, in the future, subject searching may be carried out far more by "semantic web" applications than by human searchers. The effectiveness of such applications may well depend on professional indexers even more than today's human searchers do, although this is something we can only speculate on. However, if we assume that scholars "should" be reviewing a set of documents on their topic of study that represents something close to the total population, then we may conclude that contemporary scholarship, at any rate, is likely to be significantly compromised without professional indexing present in bibliographic databases, if these results were generalized. Of course, further research is needed to investigate the extent to which they are in fact generalizable, across different disciplines, types of resources, record content, controlled vocabularies and indexing practices, database sizes and interfaces.

One might hypothesize that databases covering humanities and social science literature, and those indexing particular media types, such as images and moving pictures, are likely to more need assigned indexing to address the "synonym problem" mentioned earlier.

It was noted that the AEI database does not include any author-assigned keywords, which are present (and indexed), at least for some resources, in other bibliographic databases. This may reduce the impact of professionally assigned indexing, though, often, author keywords are based on titles and abstracts already in the records.

The indexing of the document text itself (when the resource is textual) would reduce the impact of assigned in-

dexing on "absolute" recall, but in this scenario the focus, even for scholarly searching, would probably be on "practical recall," that is, the number of relevant resources at the top of result lists, before the point at which the searcher looks elsewhere or stops altogether. Whether the "practical" result of full-text indexing is more relevant resources, and a reduction in the impact of assigned indexing, is a question yet to be conclusively answered.

ATED, the thesaurus used by AEI, includes a considerable number of cross-references, although these were not indexed in the system used for the second study, thereby reducing the impact of the vocabulary, particularly given that the search interface did not provide access to ATED. On the other hand, a "search thesaurus" could be used to increase recall (either manually or automatically) without subject indexing. An interface that displayed subject facets of result sets would encourage the use of professionally assigned terms, although, again, it could also encourage the use of, for example, author-assigned keywords.

The "depth" of the indexing in a given database (or catalog) might also significantly affect its impact. The AEI indexers add identifiers for concepts more specific than the established descriptors in ATED, but there is also the question of how "extensive" their coverage is. That is, how much treatment of a topic in an article is needed for the indexer to assign the relevant descriptor(s)? In theory, this amount should equate to the threshold for the article to be judged (potentially) relevant, in relation to that topic but does this reflect the reality?

While the second study factored in a number of real-life variables not covered by the first study, it should be noted that actual scholarly literature searches tend not to be confined to a single database. Nowadays, of course, subject searching often begins on a "discovery tool" that allows for federated, concurrent searching across a range of databases, typically including a library catalog. This would no doubt find many more resources than would searching on a single database, but whether the overall proportion of resources found only through professionally assigned indexing would be significantly different is unclear. This question also needs to be researched.

Another aspect of real-life scholarship that needs to be recognized in a comprehensive analysis of the value of professional indexing is the finding of missing resources by other means, outside of searching bibliographic databases and library catalogs. For instance, scholars might follow up on references in papers, utilize citation indexes, and/or perform author and journal searches. The results of studies such as those described in this article thus have to be considered in light of all elements of practice involved in modern scholarship.

Conversely, bibliographic databases and library catalogs are not only searched for the purposes of exhaustive literature reviews. They are often searched for known resources, or for resources by known authors, for instance, and a full valuation of professional indexing and cataloging needs to take this into account. In other search contexts, when just "some" useful resources on a given topic are needed, subject indexing might nevertheless also improve results, in terms of "precision" rather than recall. In large databases and catalogs, and in federated search systems, a suitable measure of retrieval quality in real-life searching might be deemed "practical precision," just as we noted the measure of "practical recall" above. We might be interested, for example, in the "degree" of relevance, or utility, of those resources listed in the first page of results of each search. Retrieval algorithms applied by contemporary discovery tools tend to weight subject indexing quite highly, which may significantly improve outcomes even in this search context. This question also needs to be urgently researched, and could be studied, like the impact of subject indexing on recall, through before-and-after experiments.

## 8.0 Conclusion

It would appear that, for scholarly subject searching, professionally assigned subject indexing in the *Australian Education Index* improves recall to a similar extent that professional assigned subject headings improve recall in contemporary academic library catalogs. On average, for every three relevant resources that could be found without the indexing, another relevant resource could only be found with it. In other words, over a quarter of relevant resources on a given topic would be "missed" in a typical literature search; in some cases, the proportion would be considerably higher. More research is needed to determine whether such loss levels also apply to other databases, and the extent to which they might be offset by other components of the scholarly information seeking process. However, it would be reasonable to conclude from the initial studies reported here that if the AEI were to forego its subject indexing, scholarly information needs would be substantially less well met, at least in some cases and probably in many.

## References

ACER Cunningham Library. 2017a. "Australian Education Index (AEI)." https://www.acer.org/library/australian-education-index-aei

Bawden, David and Lyn Robinson. 2012. *Introduction to Information Science.* London: Facet.

Borie, Juliya, Katie MacDonald and Elise Sze. 2015. "Asserting Catalogers' Place in the 'Value of Libraries' Conversation." *Cataloging & Classification Quarterly* 53 nos. 3/4: 352-67.

Cleverdon, Cyril. 1967. "The Cranfield Tests on Index Language Devices." *Aslib Proceedings* 19: 173-94.

Gross, Tina and Arlene G. Taylor. 2005. "What Have We Got to Lose? The Effect of Controlled Vocabulary on Keyword Searching Results." *College & Research Libraries* 66: 212-30.

Gross, Tina, Arlene G. Taylor, and Daniel N. Joudrey. 2015. "Still a Lot to Lose: The Role of Controlled Vocabulary in Keyword Searching." *Cataloging & Classification Quarterly* 53, no. 1: 1-39.

Gross, Tina. 2015. "Naming and Reframing: A Taxonomy of Attacks on Knowledge Organization." *Knowledge Organization* 42: 263-8.

Hider, Philip. 2006. "Search Goal Revision in Models of Information Retrieval." *Journal of Information Science* 32: 352-61.

Hider, Philip. 2017. "A Critique of the FRBR User Tasks and their Modifications." *Cataloging & Classification Quarterly* 55, no. 2: 55-74.

Hjørland, Birger. 2014. "Classical Databases and Knowledge Organization: A Case for Boolean Retrieval and Human Decision-making during Searches." *Journal of the American Society for Information Science and Technology* 66: 1559-75.

Kim, Heesop. 2014. "Retrieval Effectiveness of Controlled and Uncontrolled Index Terms in INSPEC Database." *Malaysian Journal of Library & Information Science* 19, no. 2: 103-17.

Mann, Thomas. 2008. "The Peloponnesian War and the Future of Reference, Cataloging, and Scholarship in Research Libraries." *Journal of Library Metadata* 8: 53-100.

Muddamalle, Manikya Rao. 1998. "Natural Language versus Controlled Vocabulary in Information Retrieval: A Case Study in Soil Mechanics." *Journal of the American Society for Information Science* 49: 881-7.

Pontis, Sheila and Ann Blandford. 2015. "Understanding 'Influence': An Exploratory Study of Academics' Processes of Knowledge Construction through Iterative and Interactive Information Seeking." *Journal of the Association for Information Science and Technology* 66: 1576-93.

Rieh, Soo Young and Hong (Iris) Xie. 2006. "Analysis of Multiple Query Reformulations on the Web: The Interactive Information Retrieval Context." *Information Processing & Management* 42: 751-68.

Rowley, Jennifer. 1994. "The Controlled Versus Natural Indexing Languages Debate Revisited: A Perspective on Information Retrieval Practice and Research." *Journal of Information Science* 20: 108-19.

Savoy, Jacques. 2005. "Bibliographic Database Access Using Free-text and Controlled Vocabulary: An Evaluation." *Information Processing & Management* 41: 873-90.

Weber, Michael A., Stephanie A. Steely, and Marilou Z. Hinchcliff. 2006. "A Consortial Authority Control Project by the Keystone Library Network." *Cataloging and Classification Quarterly* 43, no. 1: 77-98.

Zhang, Pengyi and Dagobert Soergel. 2014. "Towards a Comprehensive Model of the Cognitive Process and Mechanisms of Individual Sensemaking." *Journal of the Association for Information Science and Technology* 65: 1733-56.

Zhang, Yan. 2013. "The Development of Users' Mental Models of MedlinePlus in Information Searching." *Library and Information Science Research* 35: 159-70.

**Appendix A: Retrieval Loss in the AEI Database**

| Search query | Hits in full search (*n*) | Hits in search excluding subject indexing (*n*) | Retrieval loss (%) |
|---|---|---|---|
| lesson & planning | 270 | 176 | 34.8 |
| digital & storytelling | 70 | 51 | 27.1 |
| concept map | 56 | 56 | 0.0 |
| giftedness & music | 11 | 9 | 18.2 |
| saturday & school | 49 | 38 | 22.4 |
| astronomy | 119 | 87 | 26.9 |
| middle & school & structure | 163 | 117 | 28.2 |
| free & online & articles & about & learning | 1 | 1 | 0.0 |
| physical & activity & academic & performance & children | 10 | 9 | 10.0 |
| boys & girls & learn | 64 | 63 | 1.6 |
| gender & balance | 114 | 93 | 18.4 |

| Search query | Hits in full search (*n*) | Hits in search excluding subject indexing (*n*) | Retrieval loss (%) |
|---|---|---|---|
| differentiated & instruction | 71 | 62 | 12.7 |
| nurture & students & development & through & communication & in & classroom | 3 | 3 | 0.0 |
| writing. & learning & to & teach & english & in & secondary & school | 23 | 6 | 73.9 |
| play-based & effectiveness | 8 | 2 | 75.0 |
| angry & 'and' & aggressive & children | 4 | 4 | 0.0 |
| language & cueing & systems | 4 | 4 | 0.0 |
| reading & comprehension & importance | 61 | 44 | 27.9 |
| libraries & non & english | 46 | 23 | 50.0 |
| segregation | 195 | 148 | 24.1 |
| ecosystems | 81 | 80 | 1.2 |
| training & 'and' & crisis | 44 | 38 | 13.6 |
| positive & youth & development | 158 | 92 | 41.8 |
| intelligence & classroom | 181 | 128 | 29.3 |
| assessment & large & online & distance | 29 | 15 | 48.3 |
| assessment & large & online | 189 | 129 | 31.7 |
| listening & relations & education | 20 | 13 | 35.0 |
| learning disabilities' & 'AND' & 'brain research' | 8 | 3 | 62.5 |
| neuromyths & in & education | 5 | 4 | 20.0 |
| learning & styles & 'and' & pedagogy | 28 | 21 | 25.0 |
| youth participation' | 59 | 54 | 8.5 |
| cloud & computing | 32 | 16 | 50.0 |
| parenting & skills | 271 | 111 | 59.0 |
| sensory & play | 28 | 22 | 21.4 |
| exploratory & play | 89 | 84 | 5.6 |
| group & work & with & children | 427 | 379 | 11.2 |
| home-education | 44 | 43 | 2.3 |
| foundation & style | 68 | 43 | 36.8 |
| teacher & review & pedagogy | 172 | 132 | 23.3 |
| whiteboard & video | 16 | 13 | 18.8 |
| direct & instruction | 320 | 272 | 15.0 |
| cyberbullying | 87 | 78 | 10.3 |
| transgender | 54 | 42 | 22.2 |
| flipped & learning | 34 | 32 | 5.9 |
| animal & assisted & therapy | 5 | 4 | 20.0 |
| importance & of & science & in & primary & school | 8 | 3 | 62.5 |
| first & generation & college & student | 14 | 12 | 14.3 |
| misconceptions, & primary, & science | 73 | 16 | 78.1 |
| guided & reading & learning & difficulties | 11 | 6 | 45.5 |
| sexual & assault & on & campus | 3 | 1 | 66.7 |
| positive & discipline | 308 | 238 | 22.7 |
| out & of & school & care | 145 | 123 | 15.2 |
| gender & segregation | 54 | 34 | 37.0 |
| school & based & intervention & social & work | 62 | 42 | 32.3 |
| heavy & work | 81 | 75 | 7.4 |
| theology | 250 | 217 | 13.2 |
| authentic & student & engagement | 229 | 142 | 38.0 |

| **Search query** | **Hits in full search (*n*)** | **Hits in search excluding subject indexing (*n*)** | **Retrieval loss (%)** |
|---|---|---|---|
| reading & for & pleasure | 6 | 5 | 16.7 |
| art & therapy | 60 | 33 | 45.0 |
| year & 9 & selective | 25 | 16 | 36.0 |
| personalized & learning | 26 | 18 | 30.8 |
| new & arrival & program* | 43 | 28 | 34.9 |
| individualized & learning | 67 | 45 | 32.8 |