9-2019

Measuring the value of professional indexing

Philip Hider
Charles Sturt University

Pru Mitchell
Australian Council for Educational Research (ACER)

Robert Parkes
Australian Council for Educational Research

Measuring the value of professional indexing

Philip Hider, Pru Mitchell, and Robert Parkes

Introduction. *This study provides both a quantitative estimate and qualitative analysis of the additional 'retrieval power' that professionally assigned subject indexing affords users of a typical database in the field of education.*
Method. *A full version of Informit's A+ Education database and one stripped of its subject indexing were searched by four research assistants tasked with compiling exhaustive bibliographies on forty-eight topics. The searchers were then surveyed about their use of the two database, while their bibliographies and search logs were also examined.*
Analysis. *A two-way ANOVA model was constructed to estimate the percentage of additional resources found by the searchers on the full version of the database, while the survey responses, bibliographies and logs were analysed qualitatively.*
Results. *It was estimated that the subject indexing increases the yield of relevant resources by an average of between 48.1 and 124%, across the four searchers and all possible topics. The qualitative analysis showed that the indexing and controlled vocabulary was used by searchers in different ways, and that it also provided them with more selection power.*
Conclusions. *The study provides clear evidence that professional indexing and controlled subject vocabularies can greatly enhance the performance of those tasked with scholarly literature searches.*

Introduction

Libraries and other collecting institutions continue to provide access to many of their new acquisitions through subject headings and descriptors, typically assigned by metadata specialists employed at one point or another in the supply chain. Likewise, libraries and other information agencies commonly subscribe to bibliographic databases that, in many cases, continue to be supported by the work of professional indexers, the cost of which is ultimately passed onto the subscribers. Conventionally, the application of controlled vocabularies of subject headings and descriptors by metadata specialists and professional indexers is seen as complementing the subject access provided by titles, abstracts and other elements of derived and automatic indexing (Bawden and Robinson, 2012; Rowley, 1994; Svenonius, 1986). Clearly some of the subject terms assigned by professional indexers would not already be present in the derived indexing, and as such would retrieve additional resources. It is also likely to be the case that, for a given resource, an assigned index term, particularly based on a subject thesaurus that is also used in the searching, would sometimes more precisely represent a concept the

searcher had in mind than would a derived index term. Thus, professional indexing is seen to add *some* value to a bibliographic database. However, this raises the question: how *much* value does it add?

Seldom has research attempted to quantify the added value to databases that professional indexing and cataloguing provides, but, in today's economic climate, this research is urgently needed to inform decisions around return on investment, especially in cases where content-based retrieval alone can yield impressive results, and where the costs of professional indexing are expensive relative to search technology. This article reports on the results of a project that aimed to measure the value professional indexing adds to subject searching in a typical, contemporary bibliographic database. More specifically, it focuses on the additional *retrieval power* that assigned subject indexing provides a bibliographic database, using the assumption that for scholarly literature searches, the aim is to retrieve all relevant materials available. The measure of retrieval power is thus based on counts of resources rather than dollars, and so does not directly answer the question of professional indexing's return on investment, but it does provide a clearer basis for decision-making in this regard: in order to make an informed decision about investing in professional indexing, database managers and librarians need empirical data and specific values, i.e. they need a measure.

The data collected in the study also invited qualitative analyses of reasons for, and nature of, the added value of professional indexing estimated in the quantitative component of the research. These analyses pointed to ways in which the value might be increased, as well as to its limits.

Literature review

In the field of information retrieval, the effect of subject indexing has traditionally been measured in terms of *recall* and *precision*, and variations and combinations thereof, with recall representing the proportion of *relevant* documents retrieved from a database and precision the proportion of *relevant* documents in those documents retrieved. Through the use of these measures, the ground-breaking Cranfield experiments cast into doubt the notion that indexing using artificial controlled vocabularies, including classification schemes, was more effective than indexing derived from terms found in the document itself, such as its title (Cleverdon, 1967). As computing power and statistical retrieval algorithms developed through the last decades of the twentieth century, the claims made for content-based retrieval increased, although at the same time subject thesauri were developed to provide stronger claims for controlled indexing, despite the expense involved in manually maintaining and applying the thesauri. The competing claims were often not based on direct comparison, which was (and still is) hard to do, due in large part to the different contexts in which the two approaches were designed to be used: content-based retrieval systems tended to be experimental and based on large corpuses of documents without any assigned indexing, whereas systems supported by subject thesauri tended to be *real world*, lacking in sophisticated search algorithms and options to search the content of the indexed documents. There was (and still is) potential for disagreement among claimants as to how the competing approaches should be compared: the measures of recall and precision made sense in experimental settings, but the concept of relevance on which they were based was increasingly questioned by those who studied, as well as worked, in real-life environments where searchers might well vary in their relevance judgements depending on the situation (Schamber, Eisenberg and Nilan, 1990). Direct comparisons between derived and assigned indexing that have been attempted in post-Internet years include studies by Kim (2014), Savoy (2005) and Muddamalle (1998).

In summary, the general conclusion to be drawn from the long debate about the relative merits of assigned and derived indexing, and between controlled and uncontrolled indexing language, is that there is no clear winner: one or other might be stronger in a particular setting, but ultimately the two approaches complement each other (Rowley, 1994; Svenonius, 1986). While it remains difficult to compare the approaches in operational settings, where there are so many variables at play, some

(though not many) attempts have been made to measure the value that controlled vocabularies, as assigned indexing, *add* to a given search system. Such studies are particularly useful in the contemporary search environment, with most systems now able to make use of a good deal of content in powerful ways, making the content-based approach something of a given. An early example was Fidel's (1992) investigation of the instances in which professional database searchers used search terms based on controlled and derived indexing. Another example was the study of two controlled vocabularies used in Australian school libraries, conducted by Hider and Freeman (2009), but perhaps the best known, in recent years, are the two studies carried out by Gross and Taylor (2005) and Gross, Taylor and Joudrey (2015). They defined their research question in terms of how much was *lost* by a database not including subject indexing, instead of how much was gained by including it.

In the initial study, Gross and Taylor (2005) found that without *Library of Congress Subject Headings* (LCSH), a university library's catalogue retrieved about a third fewer records for an average subject keyword search than it did when the headings were included in its records. In the follow-up study (Gross, Taylor and Joudrey, 2015), when the catalogue had been enhanced with more tables of contents, abstracts, and other derived metadata, these subject headings still accounted for about a quarter of *hits* on average. Hider (2018) replicated the Gross studies using EdResearch Online, a bibliographic database covering scholarly papers in the education field, and found a very similar percentage loss to that of the later Gross study when the assigned subject indexing, based on the *Australian Thesaurus of Education Descriptors* (ATED), was removed from the database.

Hider (2018) has discussed the limitations of the approach taken in the Gross studies, pointing out that real-life subject searches are often not conducted in isolation, particularly if they form part of a scholarly literature search. The searcher may try a number of different queries that come to mind, or as a result of words and terms encountered in the records retrieved, including those that are a product of the assigned subject indexing. Search iteration may thus both reduce and increase the impact of the subject indexing, and it is unclear what the net result would be. To explore this, Hider (2018) conducted a pilot study using the Australian Education Index (AEI), covering a superset of the records included in EdResearch Online. A research assistant was tasked with performing literature searches for twenty topics, based on the keywords used in Hider's earlier study on EdResearch Online. The assistant was first asked to search a version of Australian Education Index stripped of its assigned subject indexing, and allowed to enter as many queries as they liked, to compile a bibliography of relevant resources for each topic. They were then asked to try to find additional relevant resources for each topic on the full version of the Australian Education Index. On average, they were able to find about another third, which equated to a loss of just over a quarter of resources without the extra indexing. This result was very similar to that of the earlier studies, suggesting that the net effect of search iteration is somewhat neutral. On the other hand, the introduction of relevance judgements (on the part of the assistant) might be a confounding factor: there may have been different proportions of relevant hits in the yields from the two versions of the database.

Two particular limitations of his pilot study were noted by Hider (2018): the small sample of topics and the even smaller number of searchers (i.e. one). Although the research assistant conducting the searching was experienced at performing literature searches and had some knowledge of the field in question (education), it is likely that not all skilled literature searchers would end up with the same list of resources for each topic, even with the large amount of time that was provided. The follow-up study reported here aimed to address these limitations.

Research design

It should be noted from the outset that the research design of this study focused on *serious* and *scholarly* searching of a particular bibliographic database. Although such databases are offered primarily for academic and research library users, it is acknowledged that they are often searched more casually and less thoroughly, where the aim is not to compile a comprehensive list of resources for a given topic. In such circumstances, the *number* of relevant resources found is not so crucial, whereas the *precision* of the initial hits becomes more important. The latter can also be measured, but this study concentrates on volume, under the assumption that rigorous scholarship aims to consider *all* the relevant literature on the topic in question.

In the pilot study topics were derived from search logs, but in the main study they were based on actual literature search requests that had been received by the Cunningham Library of the Australian Council for Educational Research (ACER), which maintains the Australian Education Index. The 494 topics that had been submitted to the Library by researchers over the period 2009-2016 were considered to be representative of possible topics. From this sample frame, fifty-two topics were randomly selected for literature searches, initially to be conducted by two research assistants, A and B. The first four of these topics were used in a *warm up* exercise. After that, in the two-factor design, A and B were tasked to search for resources on the same forty-eight (remaining) topics, but for each topic they were randomly assigned different databases, either X or Y. The forty-eight topics are listed in Appendix A.

Database Y was the full version of the A+ Education database (Informit, 2019), which used the Australian Education Index to provide access to a large number of full-text articles in the field of education, though in other cases only the bibliographic data was provided. The subject indexing for the Index was applied by professional indexers employed by the Cunningham Library, using a combination of uncontrolled identifiers, and controlled terms from the *Australian Thesaurus of Education Descriptors*, a thesaurus of concepts across all levels of Australian education that has been developed and published by Australian Council for Educational Research since 1979 (Hider, Spiller, Mitchell, Parkes and Macaulay, 2016). As at February 2019, the *Australian Thesaurus of Education Descriptors* contained over 10,000 terms, around half of which were preferred terms and half references. As part of its metadata profile, A+ Education featured four subject-related fields from the Australian Education Index schema (Informit, 2019). Subject (Major) was used to describe the subject matter of the resource using terms from the *Thesaurus* (for example, *Direct instruction* or *Student stress*). The Subject (Minor) also used the *Thesaurus* and described both the research methodology used and the education levels covered by the resource's content (for instance, *Literature reviews* and *Middle school students*). The Identifier field is a searchable, scannable index containing uncontrolled terms, representing concepts identified as candidate terms by indexers but which have not yet been incorporated in the thesaurus, as well as proper names of organisations and individuals, and names of projects, tests or reports. The fourth subject field recorded Geographic subject terms using a custom vocabulary separate to the *Australian Thesaurus of Education Descriptors*. Within Informit's A+ Education, these four subject indexes could be searched independently, or together via a combined subject search. In Database Y, the full version, users could also browse on the *Australian Thesaurus of Education Descriptors* itself, as well as on an index of all the assigned subject terms. Further, there was the option of linking from a subject descriptor in the displayed record to other records containing that descriptor.

Database X consisted of the same database, with the same interface, but stripped of the assigned subject indexing and corresponding search options described above.

The two searchers, A and B, were allowed as much time as they needed to compile a bibliography of resources for each topic using database X or Y. They were asked to include all resources they judged to be potentially relevant or partially relevant. They could inspect full-text if they wished and where it

was available. The searching was conducted over the period January to April 2018. The databases were copies of the live version of A+ Education and were not added to during the period; only the searchers and principal author were given access to the databases. Neither searcher had any relationship with the Australian Council for Educational Research or the Australian Education Index. They did, however, have a working knowledge of the field of education, as well as an employment record that included considerable experience in conducting literature searches for research purposes.

Following post-hoc analysis of the statistical power produced by the sample of forty-eight topics, it was decided not to increase the number of topics, but instead to add another two searchers, C and D, who were tasked to search on the same forty-eight topics in the same way as A and B did. They conducted their searching over the period June to September 2018. Again, searchers C and D had no connection with the Australian Council for Educational Research or the Australian Education Index, but had similar levels of search skills and education knowledge, at least on paper, as had searchers A and B. All four searchers were paid for their time and clocked up over 200 hours between them, having been asked to perform exhaustive searches for each topic. At this point, given resource constraints, it was decided not to add either additional searchers or topics.

Instead, the data collection exercise was completed by carrying out an independent relevance rating exercise on a quasi-random sample from the searchers' bibliographies for the forty-eight topics. An expert in the field of education was asked to rate the degree of relevance of the third, sixth and ninth resource listed in each of searcher A's and B's bibliographies to the applicable topic, using a four-point scale from not relevant to very relevant, based on the records (which included abstracts). The ratings were compared to assess whether the searchers' overall relevance judgement was significantly different, and, more crucially, whether their relevance judgement changed as they moved from one database to the other. Later, another independent expert rated samples of searcher C's and D's records in the same way.

Following the data collection, statistical modelling (ANOVA) was applied to estimate, at a high confidence interval and within a certain margin of error, the percentage of additional relevant resources that are found, on average, because of the assigned subject indexing on the A+ Education database. T-tests were performed on the ratings data for differences in the degree of relevance of resources compiled by the searchers, and of those complied using the two databases.

A five-question survey (see Appendix B) was administered to each of the searchers when they had completed all of their assigned topics. The survey asked the searchers to reflect on their experience of searching and on the different results obtained in the different databases, and to highlight any topics they thought the extra subject terms particularly helped with. The survey questions encouraged critical feedback and invited searchers to provide specific examples where possible. There was a question about any disadvantages experienced in using database Y, as well as any advantages, and an invitation for suggestions on ways in which the full database could make more or better use of the extra subject terms. The searchers were also asked whether they needed to look at the full-text of the articles (when available) when making relevance judgements.

Finally, two of the topics were selected for in-depth review, using the search results sheets completed by each of the pairs of searchers and the corresponding search logs. This was intended to provide additional understanding of the impact of the available subject indexing, and of the differences between searchers.

Findings

The results of the quantitative and qualitative components of the study are presented below in two separate sections.

Quantitative analyses

The t-test performed on the ratings of the resources sampled from the bibliographies of searchers A and B indicated a significant difference ($p<0.01$) in what the searchers considered relevant (or partially relevant). The percentages of *not relevant* ratings were likewise quite different: 17.4% in the case of A and 29.2% in the case of B. Similarly, the t-test performed on the ratings of the resources sampled from the bibliographies of searchers C and D indicated a significant difference ($p<0.01$) in the searchers relevance judgements. The percentages of *not relevant* ratings were also different: 25.7% in the case of C and 11.8% in the case of D. These findings are not all that surprising: it has often been pointed out how subjective relevance judgements can be, especially if based on the interpretation of a few words to represent a topic (Wolfram and Olson, 2007).

However, when the ratings of the sampled resources found on database X and Y were compared, no statistical difference was apparent: the null hypotheses for the two t-tests were $p=0.326$ (searcher A), $p=0.280$ (searcher B), $p=0.678$ (searcher C), $p=0.875$ (searcher D). With sample sizes of 72, this result strongly suggests that the searchers did not particularly lower or raise their relevance threshold as they moved between databases, an important finding given that an inconsistency would cast some doubt on the accuracy of the estimation of the additional yield produced by database Y.

Figure 1 shows the mean numbers of resources found on the two databases for each of the four searchers, while Table 1 provides summary statistics about the searchers' yield on X and Y. There are clearly significant differences between the four searchers' yields, as well as differences between the differences in their yields across the two databases. For example, searcher C found about twice or more relevant resources on database X and about three times or more resources on database Y than the other three searchers did. On the other hand, the independent relevance ratings reported above suggests that searcher C is not necessarily two or three times as effective a searcher as the A, B and D, but that some part of C's higher yields could be due to his lower relevance threshold.
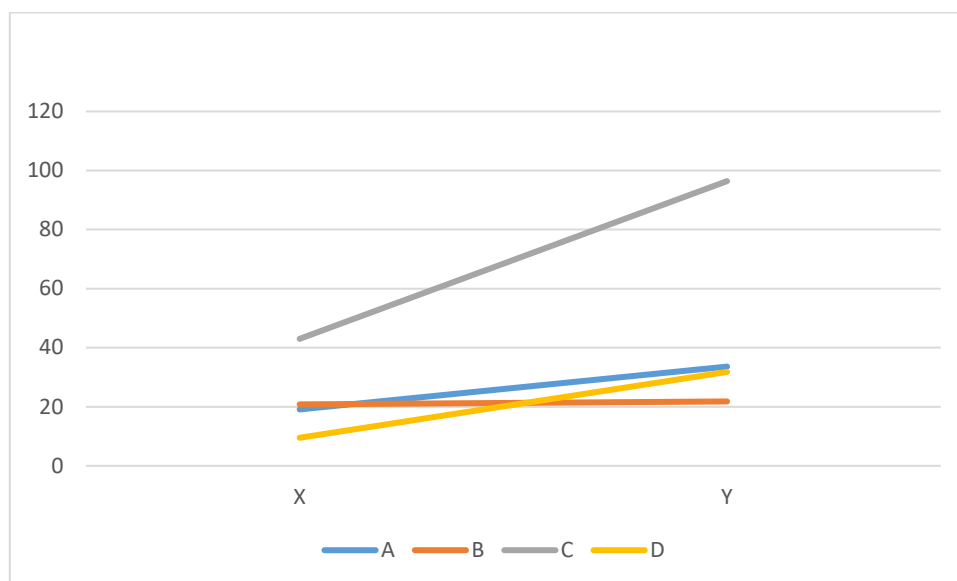
Figure 1. Mean numbers of resources found on databases X and Y by searchers A-D

	Minimum yield		Mean yield		Median yield		Maximum yield	
Searcher	X	Y	X	Y	X	Y	X	Y
A	4	9	19.1	33.6	19	29.5	40	90
B	9	7	20.8	21.8	19	23	43	35
C	4	11	43	96.4	33	82.5	131	247
D	1	2	9.5	31.8	10	19.5	21	146

Table 1. Searchers' yields on databases X and Y

Each searcher's increased yield using database Y cannot, however, be ascribed to changes in their relevance judgement, as noted above. It is likely, then, that differences in the amount of increased yield between searchers (and these differences appear considerable) would be due to differences either in their respective capacities to find resources without the assigned subject indexing or their respective capacities to find resources using the assigned subject indexing, or both.

A two-way ANOVA model was used to predict the overall effect of the variable *database* on the combined searchers' yields (i.e. if the searchers were to continue using the databases until there were no more topics to search on). Controlling the *searcher* factor, the model indicated that database Y had a mean yield higher than that of database X by $\exp(0.600) - 1 = 0.822 = 82.2\%$, with a lower bound of 48.1% and an upper bound of 124% for the 95% confidence interval. In other words, there is a 95% chance that database Y increases yield of the four searchers on average by between 48.1% and 124%.

Of course, it may be that these four searchers are not representative of the total population of users and prospective users of A+ Education, or even of those research assistants who might carry out exhaustive literature searches on the database. The four searchers were chosen purposively, for practical reasons; however, as such, they were all well suited to being employed in real life to conduct literature searches on the database in question, and they would need to be quite atypical for the mean increase in yield across all prospective, qualified literature searchers to be outside this spread of 48.1% and 124%. In any case, we can safely conclude that at least for searchers A-D, the assigned subject indexing included in the full A+ Education database accounts for at least about a third (48/148%) of its retrieval power, and likely something in the region of half of it.

Qualitative analyses

The post-search survey, and selected search results sheets and search logs, were also analysed as part of this research, providing a richer insight into the nature of the value that the subject indexing added to the A+ Education database in the context of a particular search interface. As a by-product, a greater awareness was gained of how searchers interact with this database. The following section summarises the findings from these data sources.

The provision of subject terms in database Y often affirmed the appropriateness of terms the searchers had already selected and thereby boosted their confidence. For instance, searcher A reported that '*the thesaurus essentially served to confirm search terms I considered*'. Meanwhile, searcher B welcomed the availability of the thesaurus, because '*there were some searches that did not have a direct match to a subject term so it was necessary to look for synonyms, antonyms or even similar concepts*'. Some of the searchers reported that they depended on existing subject expertise in their interpretation of the topics for searching. It should be noted that the experimental environment did not completely reflect the usual research library scenario in which the searcher would have the opportunity to interrogate the client to refine search requirements and resolve question ambiguity. Thus, the indexing served to

provide subtopics or narrower concepts that assisted the searcher, particularly given that there was no access to a client to clarify the search question.

> *The search questions could have had a number of meanings or viewpoints, and not having the means to clarify the meanings, I tried to include search results that were both broadly on topic, as well as focused on certain key aspects of the question and nuances of meaning. (Searcher B)*

When asked whether they needed to search or browse the thesaurus, searcher A responded, '*In my opinion you MUST browse in order to analyse whether the term is the one you're looking for*'. However, another searcher indicated that the subject indexing was less valuable for broader topics that provided large sets of results: '*The larger the search results the less effective the subject terms became*' (searcher C). Likewise, other searchers noted that the index terms seemed particularly valuable for very specific searches that produced small results sets. An increase from three hits to four, as in the example of Indigenous information and communication technology (ICT) ownership, could make all the difference.

A review of the search interface of A+ Education was not a focus of this research. However, it seems that in completing the survey, searchers found it difficult to separate the functionality of the system and its interface from the content. Thus, they provided some valuable feedback on the database's design for its hosts to consider. Advanced search functionality was important to these searchers. They reported on the use of Boolean operators. Searcher C was very strong on this, particularly the use of the NOT operator, and believed the '*Boolean search structure could be more reliable and more cogent and efficacious*'. The searchers expressed a desire for the option to sort results by alphabetical order, as well as by relevance. Initially this seemed surprising given that they had been tasked to find relevant resources. However, on review it was found to be an important feature, in order to enable de-duplication when the searchers were compiling their bibliographies. One searcher went through the entire project wanting to use search history functionality '*where you have the ability to add searches together, e.g. S1 AND S2*'. This functionality existed in the system being used, in fact, but when this searcher went looking for it, they were unable to locate it. Functionality that is not obvious even to an expert searcher is a waste of both functionality and searcher time. The location of the thesaurus was seen as important, and one respondent found it annoying that this advanced feature was hidden away: '*It's absolutely fine that it's not included in the basic search interface, but in the advanced search tab it was well out of the way; annoying*' (searcher A). This advice was succinctly articulated as to *hero the subject terms*.

> *In order to showcase the subject terms, why not have them appear on the results page, rather than only on the complete record? This probably crosses into marketing territory, but if the extra subject terms are the point of difference between one database and another, why not put them where people can see them? (Searcher A)*

The search logs that were analysed could inform improvements to indexing, search functionality, and the search interface of A+ Education (and potentially other bibliographic services). Initial analysis for the selected topics showed that different expert searchers searched in different ways. While some started by constructing their search using thesaurus terms when available in the interface, others commenced with their own keyword search and then used thesaurus terms that occurred frequently in the results set, clicking on the terms' hyperlinks provided in full record displays, to refocus their search. This finding has ramifications for those who teach searching, as well as for those responsible for designing search interfaces.

The first topic reviewed in-depth using the search results sheets and logs was *direct instruction of spelling*. *Spelling teaching* is an *Australian Thesaurus of Education Descriptors* (ATED) concept which has been in the thesaurus since 1984. It has the broader term, *Teaching process*. *Direct*

instruction is also a thesaurus term, sitting as a narrower term of *Teaching methods*. It was added to the thesaurus as a preferred term in 2012, with a scope note that reads,

> *Instructional method where the principal content to be learned is presented directly to the learner by an instructor. Contrasts with approaches such as discovery learning, where the content is independently discovered by the learner. Direct instruction is often characterised by carefully structured lessons, clearly defined teaching activities and constant student evaluation in order to provide feedback and corrective instruction.*

Searcher A identified thirty-one results as relevant to this topic, while searcher D identified twenty results for the same topic. Of the total forty records listed between the two searchers, 34% of the records were thus the same. Searcher D selected seven unique records, and searcher A included eighteen records that searcher D did not. When searcher A's eighteen records were reviewed for relevance, each one was found to be missing some element of specificity required for the top relevance rating. Searchers using an exact phrase search in stripped database X could find relevant resources due to the presence of these phrases in some of the abstracts. Highly relevant documents were also missed because the searcher used an exact phrase search. This case revealed several issues related to relevance, including the value of indexing in post-search filtering. The topic as provided to the searchers was not specific about the educational level. The results included articles about spelling instruction for primary school students, and for pre-service teachers, that is university level students, on how to teach spelling. Without the minor descriptors indicating educational level, the searcher in the stripped database would have to read the abstract to determine the article's relevance, or in some cases deduce this from the title of the journal, for example, *Practically Primary*. A number of the results were specific to special education or children with learning disabilities, rather than general primary or secondary education. Again, without Subject (Minor) terms, this information can only be gleaned from the abstract or the title of the journal or conference proceedings, for example, *Australian Journal of Learning Disabilities*, or *Special Education Perspectives. English (Second Language)* is another specific audience that researchers may wish to filter on when conducting a search related to spelling instruction. The variety of alternate terms for *Instruction* found in titles and abstracts in this small sample included: teaching, training, and teaching methods, while words used for *Direct instruction*, included instructivist, and direct teaching principles. Several results returned by the searcher in the stripped version of the database were about direct instruction in reading. These mentioned spelling or the related term, *Phonemic skills*. One article referred to literacy instruction which incorporated spelling. Another was about the English curriculum (of which spelling is a strand).

The second search topic reviewed was *evaluation of primary and secondary student report cards*. Using the stripped database X, searcher B found eight records of relevance to this topic. Three of these articles included the word report or reporting in their title, and all included report card, or school reports in the abstract. The relevance of searcher B's results was mixed, with two being highly relevant and most having a very specific focus on one aspect of student reports, for example reporting on students with ADHD. One abstract included the phrase '*the report card of the education department*' in an article that was clearly unrelated to student report cards. On database Y with subject indexing, searcher A was able to retrieve fourteen relevant results. Only one of these results was also in the list of searcher B's articles. The Subject (Major) used predominantly by searcher A was *Reporting (Student achievement)* which has the see reference, *Report cards*. The scope note for *Reporting (Student achievement)* was '*The communication of knowledge gained from the assessment of student learning*'. The thesaurus lists ten related terms for this concept, and several of these were used by searcher A, including *Evaluation methods*; *Grades (Scholastic)*; *Parent student relationship*; and *Student assessment*.

Discussion and conclusions

This research indicates that professionally assigned subject indexing, based on a controlled vocabulary, can increase the retrieval power of a bibliographic database by at least 50%, and likely about 100%, for at least some searchers. Most users of bibliographic databases and library catalogues would probably feel hard done by if they were asked to do without such a large number of relevant resources, at least in the context of scholarly and systematic searching. Although this added value still needs to be weighed against the particular costs associated with professional indexing, and compared with the benefits and costs of competing products and services, these sorts of numbers do provide a strong argument for database providers to think twice before jettisoning their subject thesauri and professional indexing service, especially if their search technology is no more advanced than that found on today's standard platforms.

There is a wide range of search contexts, as well as bibliographic databases, and it should be emphasised that this study focuses on the context of serious literature searching, where comprehensiveness is the aim and where time is less of a consideration. As some of the research assistants pointed out, in real life time is always *something* of a factor, even if less of one in well-resourced research projects. The research assistant whose yields increased the most using the full database also spent the most additional time using it. The trade-off between extra resources and extra time also needs to be considered when assessing the value of controlled vocabularies and professional indexing.

On the other hand, the research assistants and some of the search behaviour revealed in the case studies also pointed to the additional *selection power*, to use Warner's term (2010), that the controlled indexing gave them. This power enables searchers to define and redefine their search goals and information needs as they progress through their search session and encounter various resources and feedback from the system (Warner, 2010). The assistants felt more confident about their choices of term and resource selection, when they were confirmed by the professional indexers' subject descriptors; this encouraged the assistants to perform not only more searches, but also more sophisticated ones, using Boolean operators, for example, and refining their searches with the aid of the thesaurus's reference structure. Thus, the extra time that the use of a controlled vocabulary might require not only tends to increase the number of resources, but it may also improve their situated relevance and quality. Unfortunately, while measures of relevance and quality can be obtained from users, the degree of control that they might enjoy in their search journey would appear quite hard to measure quantitatively, in the positivist paradigm of this present study. Likewise, in the world of linked data, subject indexing based on controlled vocabularies can have an additional value outside of that provided to human searchers of the database (Szostak, Scharnhorst, Beek and Smiraglia, 2018), but what this value might be is probably even harder to quantify, and would in any case still be more a *potential* value than a realized one.

Research into the effect of professional indexing on the precision of result sets is especially needed in the case of more casual searching of bibliographic databases, when comprehensiveness is not so much the aim and when time is a particularly important factor. More specifically, the precision of the first page or two of results from a limited number of searches on a given topic is what most needs to be measured, to correspond to what is typically *found* (i.e. looked at) in this sort of searching.

Another variable that was highlighted by the both the quantitative and qualitative components of this study was the variation in the styles of the individual searchers. They used the same search interface, and the subject indexing in particular, in quite different ways. This undoubtedly had an effect on their yields and on the value of the professional indexing, though exactly what this effect was, overall, is hard to pinpoint: their styles consisted of a large number of inter-related preferences which also depended on the search situation (for example, if they were finding few or many resources). Search styles might also have more or less effect on the use and value of professional indexing depending on

whether the search is casual or serious, and depending on the design of the search interface. Irrespective of search style, however, it is recommended that if a database is to provide controlled subject indexing, this feature should be championed by the interface design: the research assistants were of accord that a hidden subject thesaurus does a search interface a major disservice.

The study also showed that the value of the subject indexing varied according to the topic, with narrower topics tending to benefit more from the descriptors. This is probably due primarily to the greater precision of more specific terms, though it could also be due to the propensity of indexers to prefer more specific terms over broader terms. There might also be some topic branches of the thesaurus used in which descriptors are more sharply defined and where there is greater coverage. Again this is likely to increase yields.

It should also be noted, of course, that not all professional indexing and controlled subject vocabularies are equal, nor are their disciplinary and user contexts the same. Bibliographic databases covering the social sciences, such as education, may be more in need of controlled vocabularies than those covering the natural sciences, for instance. The size of a database, and the kinds of documents it covers (books, papers, audio-visual materials, etc.) may also be factors. Likewise, the indexing may differ markedly in terms of the size and depth of the vocabulary it uses, and its exhaustiveness (the number of descriptors used for each document).

Finally, another consideration in real life search settings is the availability of multiple databases, with different points of access to records of the same resources, including, for example, citation databases, and databases with different controlled vocabularies. Federated search systems may also speed up searching, especially in the case of casual searching. There are known issues with discovery services being configured to preference larger, global and full-text databases in search results, regardless of relevance, and hiding smaller, specialist or index-only services (Tay, 2016). It is hard to judge what impact the wider environment may have on the value of the professional indexing in an individual database. Clearly some of the additional resources would be found by other means, but this is offset by the amount of time these other means take.

While no claim is made that the same degree of retrieval power provided by the subject indexing in the A+ Education database is necessarily to be found elsewhere, the database chosen for this research may nevertheless be regarded as a fairly typical example of a bibliographic database offering professionally assigned subject indexing, itself based on a fairly typical subject thesaurus. There seems no reason to believe that the four research assistants employed in the study were all outliers in the population of the database's searchers, and thus little reason to doubt that the results of this research point to a significant loss for scholarship and research, if professional indexing is abandoned in the reality of today's search environment.

About the authors

Philip Hider is professor and head of the School of Information Studies at Charles Sturt University (CSU), Australia. He received his PhD from City University, London and has worked at CSU since 2003. His research interests centre mainly around information organisation and LIS education. The second edition of his text, *Information Resource Description,* was published in 2018. He can be contacted at phider@csu.edu.au

Pru Mitchell is Manager Information Services at the Australian Council for Educational Research, in Camberwell, Victoria. She holds M.Ed. from Charles Sturt University, and her research interests are in controlled vocabularies and user education.

Robert Parkes is Librarian, managing the Australian Education Index at the Australian Council for Educational Research, in Camberwell, Victoria.

References

Bawden, D. & Robinson, L. (2012). *Introduction to information science.* London: Facet.

Cleverdon, C. (1967). The Cranfield tests on index language devices. *Aslib Proceedings, 19*(6), 173-194.

Fidel, R. (1992). Who needs controlled vocabulary? *Special Libraries, 83*(1), 1-9.

Gross, T. & Taylor, A. G. (2005). What have we got to lose? The effect of controlled vocabulary on keyword searching results. *College & Research Libraries, 66*(3) 212-230.

Gross, T., Taylor, A. G., & Joudrey, D. N. (2015). Still a lot to lose: the role of controlled vocabulary in keyword searching. *Cataloging & Classification Quarterly, 53*(1):1-39.

Hider, P. (2018). The search value added by professional indexing to a bibliographic database. *Knowledge Organization, 45*(1), 23-32.

Hider, P. & Freeman, A. (2009). A comparison of ScOT and SCISSH as subject retrieval aids in school library catalogues. *Access, 23*(4), 14-22.

Hider, P., Spiller, B., Mitchell, P., Parkes, R., & Macaulay, R. (2016). Enhancing a subject vocabulary for Australian education. *The Indexer, 34*(1), 25-33.

Informit. (2019). *Informit A+ Education database field information.* Retrieved from https://www.informit.org/informit-education

Kim, H. (2014). Retrieval effectiveness of controlled and uncontrolled index terms in INSPEC database. *Malaysian Journal of Library & Information Science, 19*(2), 103-117.

Muddamalle, M. (1998). Natural language versus controlled vocabulary in information retrieval: a case study in soil mechanics. *Journal of the American Society for Information Science, 49*(10), 881-887.

Rowley, J. (1994). The controlled versus natural indexing languages debate revisited: a perspective on information retrieval practice and research. *Journal of Information Science, 20*(2), 108-119.

Savoy, J. (2005). Bibliographic database access using free-text and controlled vocabulary: an evaluation. *Information Processing & Management, 41*(4), 873-890.

Schamber, L., Eisenberg, M. & Nilan, M. (1990). A re-examination of relevance: toward a dynamic, situational definition. *Information Processing and Management, 26*(6), 755-776.

Svenonius, E. (1986). Unanswered questions in the design of controlled vocabularies. *Journal of the American Society for Information Science, 37*(5), 331-40.

Szostak, R., Scharnhorst, A., Beek, W., & Smiraglia, R. (2018). *Connecting KOSs and the LOD cloud.* Paper presented at *ISKO 2018,* Porto, Portugal. Baden-Baden: Ergon.

Tay, A. (2016). Managing volume in discovery systems. In L. Spiteri (Ed.), *Managing metadata in Web-scale discovery systems* (pp. 113-135). London: Facet Publishing.

Warner, J. (2010) *Human information retrieval.* Cambridge, MA: MIT Press.

Wolfram, D. & Olson, H. (2007). Method for comparing large scale inter-indexer consistency using IR modeling. *Proceedings of the 35th Annual Conference of CAIS–Actes du congrès annuel de l'acsi.* Retrieved from https://journals.library.ualberta.ca/ojs.cais-acsi.ca/index.php/cais-asci/article/view/250/212

Acknowledgements

Appendix A. Topics searched

Survey of graduates five years after graduation

Test validity in students from cross cultural differences

Educational resources and Mathematics achievement in junior secondary school

Test accommodations for students with learning disabilities

How Year 12 students do post-school in their academic studies

Consensus building and cross cultural negotiation

Evaluation of national and international assessment programs

Multiple choice questions and distractors

Direct instruction of spelling

Teaching and Learning School Improvement Frameworks

Evaluation of new Teacher Quality National Partnership (TQNP) programs

Cultural awareness and relevance in early childhood curriculum

Trends in online and distance learning

The role of feedback on intelligence, listening, language

Attendance and non-attendance at school

Indigenous ICT ownership

Training of secondary school mathematics teachers

Aboriginal education - equality, disadvantaged, policy, programs in primary and secondary schools

Nonverbal and verbal low-functioning children with Autism

Forms of literacy associated with reading and interpreting mathematical texts

Review of frameworks in vocational education systems worldwide

Parent, student and community attitudes to education

Models of science teaching in classrooms

Service learning in secondary schools

Emergency hazards and programs in schools

Trends in journal publishing and library collection management of journals

Engagement and retention of secondary school students

Supply and demand for Higher Degree Research qualifications in Australia

Trends and statistics of vocational education

Problem solving for secondary and higher education students

Indigenous education databases

Directing resources to schools with the greatest needs

Assessing spelling and the use of editing or dictation as the mode of delivery

Prerequisites to admission to undergraduate education courses in Australian universities

Student centred learning in secondary schools

Educational leadership in secondary schools

Parent attitudes to education

ICT and student engagement in university and higher and adult education

University graduates employment potential

Beginning teacher induction

Early childhood education for indigenous children

Financial literacy for the Australian Indigenous population

Integration aides and teacher aides for children with autism and Asperger's

Oral language and verbal ability of early childhood and primary students

Detection of cheating on standardised tests
Early childhood literacy education
Evaluation of primary and secondary student report cards
Socioeconomic background and student wellbeing

Appendix B. Survey questions

(1) The results show a significant increase, on average, in the number of relevant articles you found using the second interface (database Y).
o   Why do you think this was?
o   Was it mainly because your keyword searches picked up additional records with the extra subject terms in, or also because you saw new terms to search/click on, when you browsed through records?
o   Did you use the thesaurus and did this help at all?
o   Or the other subject indexes?

(2) What sort of topics do you think the extra subject terms particularly helped with? (Provide examples if you can think of any)

(3) Were there any disadvantages to using the second interface?

(4) Can you think of any ways in which the second interface could make more/better use of the extra subject terms?

(5) Did you find yourself needing to look at the full text of the articles (when available), when making your relevance judgements?

How to cite this paper

Hider, P., Mitchell, P. & Parkes, R. (2019). Measuring the value of professional indexing. In *Proceedings of RAILS - Research Applications Information and Library Studies, 2018, Faculty of Information Technology, Monash University, 28-30 November 2018. Information Research, 24*(3), paper rails1808. Retrieved from http://InformationR.net/ir/24-3/rails/rails1808.html (Archived by the Internet Archive at https://web.archive.org/web/20190818105206/http://informationr.net/ir/24-3/rails/rails1808.html)