Alphabetization in indexes: experimental studies

James Hartley, Lindsey Davies and Peter Burnhill

James Hartley and Lindsey Davies are from the Psychology Department at the University of Keele; Peter Burnhill is Head of the Design Department, Stafford College of Further Education

The authors have previously published work on the typography and layout of indexes. This paper reports on how students and children sequence words and retrieve items from indexes with different sequences.

A detailed reading of different texts on how to construct an index presents many contradictions for the reader. Apart from problems concerning typography and layout, issues which we have discussed elsewhere (Burnhill *et al*, 1977) one further question that receives considerable attention is the best way in which to organize the sequence of items—should it be alphabetized letter by letter or word by word? The British Standards Document BS 3700 (1976) has this to say on the matter:

> *6.1 General.* Alphabetical arrangement is generally the most practical, and is therefore the commonest, method of displaying indexes to publications. If any other arrangement is seen unquestionably to be more suitable for a particular work, the arrangement chosen should be clearly stated, and if necessary, explained. A dynastic index, for example, may be arranged *chronologically*; however, it will almost certainly need to be supplemented by an alphabetical list of rulers' names, either as a separate list or as part of the work's general index.

> Within an alphabetical index, subheadings may occasionally be more suitably arranged by a method other than alphabetical.

> *6.2.3.1. Letter-by-letter and word-by-word alphabetization.* Compound headings of two or more words may be treated as single entities alphabetized throughout (letter-by-letter), or as groups of separate words each alphabetized in turn (word-by-word). Hyphen and oblique strokes are normally treated as spaces for the purposes of word-by-word alphabetization.

> Letter-by-letter alphabetization has the advantage that a given compound heading always occupies the same position whether shown as two or more separate words (e.g. Water glass), as a hyphenated compound (Water-glass), or as a single word (Waterglass).

> Word-by-word alphabetization, however, sometimes results in clearer groupings of related

entries, since the position of any compound heading depends on whether its word elements are separated or joined.

The following examples illustrate the two methods.

Letter-by-letter alphabetization	Word-by-word alphabetization
Newark	NSW
Newcastle (NSW, Australia)	New Castle (Pa., USA)
	New Haven (Conn., USA)
New Castle (Pa., USA)	New Testament
Newels	New York
New Haven (Conn., USA)	Newark
Newhaven (England)	Newcastle (NSW, Australia)
New Testament	
Newton	Newels
New York	Newhaven (England)
NSW	Newton

Stockerau	Stock Exchange
Stock Exchange	Stock market
Stockholm	Stock-room
Stock market	Stockerau
Stockport	Stockholm
Stock-room	Stockport

Clearly then the different methods have different advantages. In the example below the letter-by-letter technique seems best, for the two concepts (fundamentalism and funds) are not split by this procedure:

Letter-by-letter	Word-by-word
fundamentalism	fund raising
fundamental theology	fundamental theology
fund raising	fundamentalism
funds	funds

However, in the example below, the word-by-word techique seems superior for, in this case, the two related words (Highway accidents, Highway regulations) are not split up:

Letter-by-letter	Word-by-word
Highway accidents	Highway accidents
Highwaymen	Highway regulations
Highway regulations	Highwaymen

So far all that we have said is common knowledge, but, so far as we have been able to ascertain, there has been little research conducted on how the *users* of indexes respond to the different sorts of alphabetical

arrangements, or indeed how non-specialists might order words themselves. A letter by Bakewell (1966) suggests that library clerical staff in industrial libraries prefer to sort catalogue cards letter-by-letter, and Ayres (1979) has reported on a study carried out on library filing cards by Thompson (1976) in an Australian College. Thompson showed that two thirds of first-year university students sorted the title cards using the method of word-by-word, but that by the third year the students were evenly divided between the two methods. The experiments to be reported in this paper can thus be considered as additional beginning steps in the direction of obtaining some factual evidence on these issues. Three pilot studies are first discussed in order to indicate how non-specialists order words: these studies are then followed by a main experiment designed to assess the effects of different sequences on the retrieval of items from an index.

Pilot Study 1: Do students order words in a consistent manner?

A group of 23 university students at the University of Keele were given the following sets of words and asked to 'put them in order for an index':

character	music	reading
characterization	musicality	ready
character-assassination	music-case	readiness
character-study	music-box	reading age
characteristic	musician	reading ability

What we hoped to see from this enquiry was whether the words would be grouped letter-by-letter, word-by-word, or by some other method, and whether or not the students would be consistent for all three sets of words. The results were quite remarkable and can be listed briefly as follows:
- only seven students appeared to attempt a consistent alphabetical arrangement;
- six of these students used the letter-by-letter and one the word-by-word arrangement;
- most of the students arranged the words by meaning, but many different arrangements appeared.

Of course it was not clear whether these results were due to our providing such short lists of words. Rearranging only five words would not reveal to the students the gross difficulties that such non-systematic approaches would produce with longer lists. Nor was it clear whether the students were actually trying to order words for how they thought indexes should be produced (i.e. a new way), or whether they thought that this was how they actually were constructed. These considerations led to Pilot Study 2.

Pilot Study 2: Further student data

Twenty-seven university students were given the following (longer) set of words and asked to write out

the words 'in the order you would expect to find them in a book's index'. After completing this task they were then asked to explain the reasoning behind their solution.

charitable
character
chain reactions
charmed life
characterization
chain mail
chains
character assassination
chain stores
charm school
character study
charity

The results of this study were more clearcut:
- Twelve students (i.e. approximately half the group) used a concept grouping. This was usually done in two forms: either key words first in a basically alphabetical order (e.g. character, character assassination) or single words before double entries (e.g. character, characterization, character assassination).
- Eight students used a letter-by-letter arrangement (one of them incorrectly).
- Four students used a word-by-word arrangement (one of them incorrectly).
- The remaining three students used idiosyncratic methods (two starting with character) and all with alphabetical errors.

Thus grouping by meaning was the most common method and approximately 20% of the sample made alphabetical errors. A typical example is provided here:

chains
chain mail
chain reactions
chain stores
character
characterisation ·
character assassination
character study
charity
charitable
charmed life
charm school

Pilot Study 3: Data from schoolchildren

One hundred and nine-five pupils (101 boys, 94 girls, aged 12-13 years) were given the same set of words as in Pilot Study 2 and asked to 'put them in order for an index'. These children came from eight mixed-ability classes in a comprehensive school.

The results of this study were also clearcut:

E

East Anglia 41
East End 83
East Indies 94
Eastbourne 12
Eastern churches 52
Education Act 56, 82
    statistics 90, 141
Educational supplies 144
Electric bells 27, 29
    lighting 30
Electrical energy 51
Electricians 61-6, 70
Electricity 71-80, 96, 101
Electrometallurgy 21
Electron miscroscope 30-60
Electronics 20-36, 42, 151, 180, 190-201
Europe, history 54
European Wars 62-66
Express service 70-100
Expressionism 110-120

Figure 1 (a). Word-by-word arrangement

E

East Anglia 41
Eastbourne 12
East End 83
Eastern churches 52
East Indies 94
Education Act 56, 82
Educational supplies 144
Education statistics 90, 141
Electrical energy 51
Electric bells 27, 29
Electricians 61-6, 70
Electricity 71-80, 96, 101
Electric lighting 30
Electrometallurgy 21
Electronics 20-36, 42, 151, 180, 190-201
Electron miscroscope 30-60
European Wars 62-66
Europe, history 54
Expressionism 110, 120
Express service 70-100

Figure 1 (b). Letter-by-letter arrangement

- One hundred and thirty-five pupils (69%) used a letter-by-letter sequence and were either correct or made one error.
- Thirty-one (16%) used a letter-by-letter sequence, but made several errors.
- Three pupils (1.5%) grouped words by sense, but the ordering between groups and within groups was not alphabetical.
- The remainder (13%) did not complete the task or made so many errors that no clear method of organization was apparent.

In short 85% of the pupils ordered the words letter-by-letter, 1.5% grouped by sense, and none used the word-by-word method of sequencing. About 30% of the group made alphabetical errors of some kind.

An experimental comparison study

The results of the pilot studies described above suggest that different methods of approach are used by different people when they are asked to put words in order for an index: students seem more likely to group by meaning and children to use a letter-by-letter approach. The next question we asked was whether or not the way in which *an index is conventionally sequenced* affects the ease of item retrieval.

A typical approach taken by research workers, when concerned about such an issue, is to make the issue more simple and to test it in controlled conditions—in order to isolate the important variables. We adopted this approach in order to see from which system it would be most easy to retrieve items. We constructed a fictitious

index by taking words from *Sears' list of subject headings* (Frick, 6th edition, 1950) and the *Concise Oxford dictionary*. Our purpose was to construct an index which would exaggerate the differences between a letter-by-letter and a word-by-word arrangement. Most of the items chosen for inclusion led to a different sequence in one or other of these two systems. A third version, which we called a 'concept' arrangement, was constructed in such a way that related word-groupings would not be split and that 'key words' would head the list of related entries regardless of their spelling. The aim of the experiment to be reported, therefore, was to see how easily schoolchildren would be able to retrieve items from these three systems of arrangement, bearing in mind (i) how they order words themselves, and (ii) that the first two systems have strong advocates in textbooks in indexing, and that the third system—to some extent a hybrid—has been considered by some to be 'disastrous' (e.g. Collison, 1972, p.89). Extracts from the three versions of the index are provided in figure 1 for illustrative purposes. It is important to bear in mind here that these are *fictitious indexes designed to exaggerate the differences in item sequence between them.*

For each version of the index we kept a number of subsidiary typographic variables constant. These were:
Alphabetic key letters ranged left.
Two line spaces between key letter and first entry.
Three line spaces between last entry under one heading letter and next alphabetic key letter.
Indent of three letter spaces for sub-items.
Indent of six letter spaces for carry-over lines (both

E

Figure 1 (c). 'Concept' arrangement

headings and sub-headings).
Capital letter for first letter of the main entries.
Lower-case for sub-items.
Colons to indicate referral to another entry.
Commas between page references, but not between entry and first page reference.
Page references abbreviated.
Page references listed in numerical order.
All entries printed with the same size type and typeface.
No full stop after entries.

The three versions of the index were typed out on A4 paper with a maximum line lenth of 35 elite characters per line and a line feed of 1½ units. These materials were then reduced 25% in size by xerox copying—this procedure producing a text image equivalent to 10 point type on a 12 point body and a column width of 13 ems (55 mm)—i.e. an appropriate size for a two-column structure on an A5 page.

A list of target words and an appropriately worded instruction sheet was attached to each style of index. The instructions read as follows:

This booklet contains an index to a textbook. Your job is to find and circle the page numbers for the items listed on the extra sheet provided as quickly as you can. Please find the items in the order listed, circle the page numbers on the index, and tick off each item on the list as you do it.

In order to create the list of target words we selected items which appeared in different positions relative to other items in at least two of the indexes. We attempted to achieve an even distribution of these items throughout the indexes. The items were then re-ordered at random—with the proviso that items close together in an index were not placed consecutively in the target list, and in addition we tried to ensure as much page turning as possible.

The instruction sheet was followed by an extract from an index printed in the same sort of sequence that the reader was to work with, headed *Example*. Three items were used to demonstrate the principle of finding an item, circling the appropriate page number and ticking off that item. The readers were then given a period of six minutes to find and circle as many items in their index as they could.

One hundred and seventy children (aged 11-12 years) acted as participants in this study. These children were first-year pupils of varying ability drawn from six unstreamed classes in a local comprehensive school. (Our earlier investigations (Burnhill *et al*, 1977) had shown that children of this age were capable of carrying out the search task which we envisaged.) Each class of pupils was taken individually, and the children were allocated at random into one of the three conditions—to search indexes in the letter-by-letter sequence, the word-by-word sequence, or the 'concept' arrangement.

Results

The results obtained are shown in Table 1. Analysis of variance showed that there were no significant differences between the average scores obtained on each of the three sequences but that girls performed significantly better than boys ($p < .05$).

Table 1.   Results of Experiment 1: The average number of items correctly located in six minutes

		Letter-by-letter sequence	Word-by-word sequence	Concept sequence
Boys	Average score	22	23	21
	Standard deviation	7	6	5
	Number of pupils	21	18	16
Girls	Average score	25	23	25
	Standard deviation	6	6	7
	Number of pupils	25	30	30

Concluding comments

The pilot studies reported in this paper have indicated that students use a variety of methods when asked to put words in order for an index—with sense and meaning being a prime consideration—but that schoolchildren are likely to be less ambitious in this respect and to order words alphabetically using the letter-by-letter method. Our main experimental study—using fictitious indexes to exaggerate the differences between sequences, and employing schoolchildren as searchers—suggested that no one method of sequencing indexes led to better retrieval than any of the others that we used. However, as the letter-by-letter and word-by-word methods of constructing indexes are considerably easier than the concept method (which is highly subjective), it would appear that either of these two methods, currently used by indexers, can be employed without causing undue difficulty for young readers.

Acknowledgements

References

Ayres, F., It's not as easy as ABC. *Catalogue and Index,* (54 Autumn 1979, 1-3, 8.

Bakewell. K. G. B. Letter to the editor. *Catalogue and Index. (4) Oct. 1966, 5.*

British Standard 3700: Recommendations for the preparation of indexes for books, periodicals and other publications. British Standards Institution, 2 Park Street, London W1, 1976.

Burnhill, P., Hartley, J. and Davies, L. Typographic decision making: the layout of indexes. *Applied Ergonomics,* **8** (1) 1977, 35-39.

Collison, R. L. *Indexes and indexing.* (4th edition) London: Benn, 1972.

Frick, B. M. *Sears' list of subject headings.* (6th edition) New York: H. W. Wilson & Co., 1950.

Thompson, M. The letter of the lore. *Australian Library Journal,* **25** (6) 1976, 196-198.

The amateur indexer/
and the local historian

Catherine D. Linehan

Sidmouth is fortunate in having a wealth of documented history; fortunate too in having its own 19th-century historian, Peter Orlando Hutchinson, who came to Sidmouth as a boy of 15. His father, Andrew Hutchinson, M.D., M.R.C.P., F.R.S., brought his family here in 1825, and practised in the town until his death in 1846. P. O. H., as our historian is known, studied as an architect but never worked as such; he spent his life seeking out and recording local history, from the earliest geological periods in which the fertile Sid Valley was formed until a few years before his death in 1897.

The result was a five-volume leather-bound manuscript history, charmingly illustrated with sketches and drawings; maps and plans abound, and all the hillforts, earthworks, and barrows are carefully surveyed and described; many of the documents from a Cartulary of about 1260 are copied in almost facsimile form, together with others from the Parish chest, feoffees' charity deeds, and so on. The pages are interleaved with letters from contemporary antiquarians—mainly local clergy—and there is a series of etchings of the houses of the neighbouring gentry, mostly built in the late 18th and early 19th centuries. All this, but alas! no satisfactory index.

The History was bequeathed by P.O.H. to the Exeter Museum, as Sidmouth had no permanent museum before the 1939-45 war. It is loaned to Sidmouth for the summer, when its museum is open; anyone wishing to consult it, though, has to work from 10 volumes of a photocopy made some years ago. The lack of a valid index to the work made this difficult.

Dr Gerald Gibbens, Curator of the Sidmouth Museum, stressed the need for an index. I had *Book indexing,* the booklet written by my sister, Margaret Anderson, and I offered tentatively to attempt the work under her guidance. I began the work in the autumn of 1979, seated at a large table with the 10 photo-copy volumes laid out in order, a shoe-box for filing, and packets of 5 in. by 3 in. cards. It was a fascinating task; I had long wanted to read this history, and the close study needed for the index made the reading doubly worth while.

More a digest than an index?

As I read, I searched for P. O. H., the man, in his writings; but he remained elusive, even when I had got used to his long ss and somewhat faded script. He had clearly had a normal active boyhood, enjoyed the hospitality of many of the local families, and taken an active part in the institutions and charities of the town.