Jointly published by Akadémiai Kiadó, Budapest and Springer, Dordrecht

Scientometrics, Vol. 63, No. 2 (2005) 341–356

What's in a title? Numbers of words and the presence of colons

GRANT LEWISON,[a] JAMES HARTLEY[b]

[a] *Department of Information Science, City University, London (UK)*
[b] *Department of Psychology, Keele University, Staffordshire, (UK)*

Much has been written about titles in scientific journal articles but little research has been carried out. We aimed to assess in two studies how factors like the length of a title and its structure might vary in different scientific fields, and whether or not these features have changed over time. Statistical analyses were made of 216,500 UK papers from science journals, and of 133,200 international oncology papers. Factors examined included title length, the use of colons in the titles, and the number of authors. All of these factors increased over time for both sets of papers, although there were some disciplinary differences in the findings. In both studies, titles with colons occurred more frequently with single than with multiple authors except when the numbers of co-authors were large. Certain features of titles can be related to different disciplines, different journals, the numbers of authors and their nationalities.

Introduction

Much has been written about the importance of titles in scientific journal articles but little research on the nature of titles has been carried out. There is plenty of exhortatory advice on how to write effective titles (DAY, 1998; O'CONNOR, 1991; HALL, 2003) but few evidence-based studies. Furthermore, until recently, the research that has been carried out on titles has been limited by the technology available to researchers at the time of their studies.

There has been some debate about the purpose of titles (HARTLEY, 2005; YITZHAKI, 2002). There are two main issues here. Titles need to attract a reader – to signal: 'here is something you need to read'. But titles also need to inform a reader: 'this is what this paper is about'. Attracted and informed, the reader can then decide whether or not to proceed further. MABE & AMIN (2002) found, for example, that 5,000 science readers estimated that they read, on average, 97 articles per year, twice as many (204) abstracts, and ten times as many (1,142) titles.

Titles thus convey important information. It is to be expected, therefore, in this age of information retrieval, that the titles of journal articles will be getting longer.

Received November 3, 2004

Address for correspondence:
GRANT LEWISON
Department of Information Science, City University, London, EC1V 0HB, UK
E-mail: g.lewison@soi.city.ac.uk

0138–9130/US $ 20.00

And, indeed this is the case, as testified by the studies reported by BERKENKOTTER & HUCKIN (1995), LEWISON (1999), LIU (2003) and WHISSELL (2004). Research has shown, perhaps unsurprisingly, that if there are more informative words in a title, then more relevant articles and fewer anomalies are likely to be retrieved in database searches (KILGOUR, 2004; KOSTOFF et al., 2004).

LEWISON & PARAJE (2004) found that they could classify biomedical research journals along a continuum from clinical (scored at 1.0) to basic research (scored at 4.0) on the basis of a selection of words used in the title of an article, the authors' addresses, and key words in these addresses. The words used in the titles were categorised as either 'clinical' (N = 121) or 'basic' (N = 107). The authors explained that these chosen words were rather arbitrary, and could be added to, but that they served their purpose of discriminating between journals. The full list of words is provided by LEWISON & PARAJE (2004): some examples of clinical words are: *abdominal, adults, American, anterior* and *assessment*, and some basic words are: *activates, adenovirus, apoptosis* and *assembly*.

One particular way of composing titles increases the number of words required and this can make them more informative. This is the form used in the title of our paper - a two-part structure - usually separated by a colon. The following example illustrates how by changing a title into a two-part format, and by adding in significant words, authors can make a title both more informative and more useful for database searches.

> Original title: *Students' perspectives on constructivist learning.*
> 1st revision: *Constructivist learning in higher education: students' perspectives.*
> 2nd revision: *Constructivist learning in higher education: postgraduates' perspectives.*
> 3rd revision: *Constructivist learning in higher education: Eight postgraduate interviews.*

DILLON (1981) noted, in the psychology journals of his time, that titles without colons contained – on average – 8 words compared to titles that included them, which averaged 17. More recently we have found that the average number of words in the titles for *Scientometrics* (1994–2003) is 11 for titles without colons and 14 for titles with them, so the presence of a colon here adds on average about three words to the length of a title.

HARTLEY (2005) summarised the findings of ten reports on the percentage of colons in the titles of research articles. These, including one of his own, were those of ANTHONY (2002), DIERS & DOWNS (1994), DILLON (1981; 1982), FONTANET et al. (1997; 1998), MICHELSON (1994), PERRY (1985) and TOWNSEND (1983). Basically it appears that there are disciplinary differences in the use of colons (with disciplines like

computer science, for example, using many fewer titles with colons than disciplines like psychology (roughly 7% vs 50%). Our studies of *Scientometrics* (1994–2003) show that 42% of the titles contained colons.

There are two other findings that might be important with respect to titles. First, it appears that the more authors there are to a particular paper, the longer their paper might be (KUTCH, 1978; YITZHAKI, 1994). Indeed, it is a common observation that the number of authors per science paper has been steadily increasing (e.g., see CRONIN, 2001; MABE & AMIN, 2002; MEADOWS, 1998) but we know of no research relating this to increases in the lengths of titles. Second, it appears that science papers with longer titles might also be longer in themselves (YITZHAKI, 2002). Putting these findings together, we might expect there to be more papers with colons in their titles today as the numbers of authors and (possibly) title lengths increase over time.

Finally, as noted above, most of the previous studies of titles and their effects have, of necessity, been restricted by the technology available to researchers at the time of their enquiries. This has meant that most (but not all) of the previous studies reviewed above have relied on manual counting in a limited number of journals and disciplines. In the two studies reported in this paper we use much larger electronic databases to assess how factors like the length of a title and its structure might vary in different scientific fields, and to see whether or not these variables have changed over time. Indeed, we carried out statistical analyses with a set of 216,500 UK science papers, and with a set of 133,200 international oncology papers. We examined changes over time in title lengths, the use of colons, and the number of authors, and we also made international comparisons.

Method

Our studies were based on substantive papers (defined as articles, notes and reviews) contained in the Science Citation Index (SCI) CD-ROM version. Two selections were made and bibliographic details (authors, titles, full source) downloaded for each of the five years 1981, 1986, 1991, 1996 and 2001. The first selection included all UK science papers in different fields. The second included all papers on cancer research from the whole world and, for these, we also downloaded their addresses. Table 1 shows the numbers involved.

For the UK science papers, each journal was assigned to one of eight major fields by means of a classification system developed by CHI Research Inc., and used for the US National Science Foundation's Science and Engineering Indicators Reports. Table 2 lists these eight major fields.

Table 1. Groups and sets of papers downloaded from the Science Citation Index for analysis, with numbers of authors per paper (A)

Year:	1981	1986	1991	1996	2001	Total
UK, all papers	33631	37924	41193	50414	53351	**216513**
A	2.62	2.99	3.34	4.50	5.05	
World, cancer	17073	21554	25680	32400	36510	**133217**
A	3.50	4.03	4.58	5.20	5.74	

Table 2. Major fields used for the classification of UK papers

Year:	1981	1986	1991	1996	2001	Total	%
Biology	3309	3661	2939	3578	3796	**17283**	8.0
Biomedical research	5356	6361	7236	8908	9167	**37028**	17.1
Chemistry	3692	3719	4235	5254	5084	**21984**	10.2
Clinical medicine	11920	14601	15929	18072	18550	**79072**	36.5
Earth and space	1694	1967	2157	3432	4234	**13484**	6.2
Engineering & technology	2357	2554	2631	3302	4296	**15140**	7.0
Mathematics	790	699	685	745	988	**3907**	1.8
Physics	3766	3700	4814	6419	6251	**24950**	11.5
Other	747	662	567	704	985	**3665**	1.7
TOTAL	**33631**	**37924**	**41193**	**50414**	**53351**	**216513**	**100**

The papers on cancer research were selected by means of a "filter"[*] based on specialist journals and title words; this process has been described earlier (LEWISON, 1999). We also conducted a geographical analysis so as to identify the outputs of the 20 largest countries. Table 3 lists these countries with their digraph ISO codes, and numbers of papers in each CD-ROM year. For this purpose, the former East and West Germany were both unified as "Germany" and papers from the former USSR were grouped with those of Russia.[**]

These international cancer papers were also classified by the research level (RL 1.0-4.0) of the journal in which they were published using the method described above. As noted, this depends upon the presence of one or more from a list of 121 "clinical" and 107 "basic" words in their titles, and their authors' addresses. Journals with only "clinical" papers were assigned to RL = 1.0; journals with only "basic" papers were assigned to RL = 4.0. All the other journals (in practice, virtually all of them) were given an intermediate value of RL.

[*] The filter "ONCOL" was developed in consultation with Dr Lynne Davies of Cancer Research UK, a leading medical charity, and had a precision (specificity) of p = 0.95 and a recall (sensitivity) of r = 0.90.

[**] The table incidentally reveals major increases in output from China and Spain, while Japan and Italy also doubled their presence in world cancer research. But Russia's output declined drastically, and the relative presence of the USA and Denmark also decreased significantly.

Table 3. List of 20 largest countries in terms of output of cancer research papers

Country	Code	1981	1986	1991	1996	2001	Total
Australia	AU	255	334	421	619	752	**2381**
Austria	AT	141	186	266	400	583	**1576**
Belgium	BE	180	246	324	475	534	**1759**
Canada	CA	540	784	981	1298	1368	**4971**
China (Peoples' Rep)	CN	65	168	219	403	801	**1656**
Denmark	DK	194	282	320	357	326	**1479**
Finland	FI	123	209	281	365	428	**1406**
France	FR	895	1080	1536	2029	2199	**7739**
Germany (incl. DDR)	DE	1236	1485	1761	2649	3473	**10604**
India	IN	144	166	213	280	295	**1098**
Israel	IL	228	327	311	416	475	**1757**
Italy	IT	534	869	1469	2269	2418	**7559**
Japan	JP	1026	1841	2734	3948	4858	**14407**
Netherlands	NL	292	582	814	1096	1194	**3978**
Russia (USSR in 81-6)	RU	451	515	624	190	173	**1953**
Spain	ES	88	132	337	667	792	**2016**
Sweden	SE	424	524	649	843	930	**3370**
Switzerland	CH	222	279	413	577	668	**2159**
United Kingdom	UK	1263	1963	2484	2855	2859	**11424**
United States	US	8263	9689	10731	12956	14363	**56002**
WORLD		**17039**	**21528**	**25654**	**32338**	**36468**	**133027**

Note: totals differ from those in Table 1 as the ones here exclude papers with no listed authors.

The papers were also classified individually as being "clinical", "mixed" or "basic", depending on whether their titles contained just a clinical word, words from both lists, or just a basic word. Not all of the papers had such a designation, but 70% overall did so. As Figure 1 shows, 52% were "purely clinical", 11% were mixed and 37% were "purely basic".* We also used this journal classification scheme for the UK papers in the fields of biomedical research and clinical medicine to allow us to make some comparisons between the data from the two sources.

* A macro, written for City University by Dr Philip Roe, was used for this purpose. A further macro, also written by Dr Roe, was used for the geographical analysis of the cancer papers.

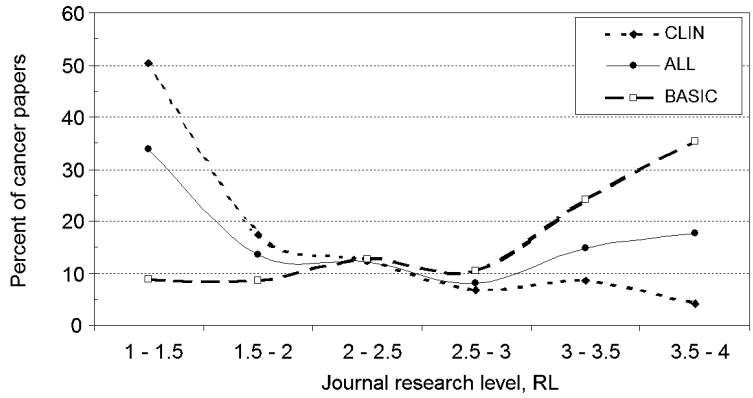

Figure 1. Distribution of world oncology papers, 2001, by journal research level, RL (1 = clinical, 4 = basic) for all papers, and those with "clinical" title words (CLIN) and those with "basic" title words (BASIC)

Analyses

The details required from the ten sets of papers shown in Table 1 were transferred to MS Excel spreadsheets for analysis. The bibliographic sources were parsed to leave just the journal name. The paper titles were processed to remove parentheses and commas, and the string "(space)-(space)" was replaced by "$" so that its presence in the title (indicative in the SCI of the presence of a colon) could be determined and marked (by means of a "1" in a column labelled C). Individual hyphens and single spaces in the title were then also replaced by "$", and the numbers of $ symbols in each modified title were counted.* Unity was added to this count to give the number of words per title, W/t.

We used multiple regression (SPSS version 11) to analyse the data since the values of the two dependent variables, namely W/t and the chance of there being a colon in the title (%C), could be influenced not only by each other but also by other independent variables.

For the UK science papers, the independent variables were:

- authors per paper, A, expressed as $\log_{10}A$ and $(\log_{10}A)^2$;
- time, T (1981 = 1, 1986 = 6, etc.);
- RL (for two fields only),

and the analysis was conducted on papers from each of the eight fields in turn.

* A macro written by Dr Judit Bar-Ilan of the Hebrew University of Jerusalem was used for this process.

In order to make a comparison between the fields, the values of W/t and %C corresponding to A = 1 and A = 4 were calculated. This enabled their variation with time to be seen independently of any changes in author number and RL that may have occurred.

For the world cancer papers, the independent variables were:

- authors per paper (on a log scale, as above);
- country (1 or 0);
- RL for each journal;
- RL for each paper (1 or 0 for "clinical" or "basic");
- time (as above),

and the analysis was conducted for two spans of years, 1981–1991 and 1996–2001,* for those papers with either a "clinical" or a "basic" word in their titles.

Results of Study 1: UK science papers in eight different fields

It was evident for the UK science papers that the fields had a significant influence on the two parameters being studied, W/t and %C. Figure 2 shows the variation of these variables for the different fields for papers with four authors, based on the regression coefficients determined by the SPSS program. The number of words per title is greatest for biology (average 15 words) and least for mathematics (8 words); colons in the titles are most evident in clinical medicine (23%) and least in engineering and technology (6%).

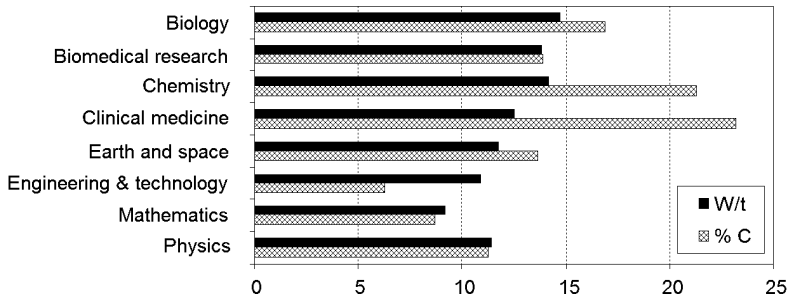

Figure 2. Variation of words per title (W/t) and percentage use of colons (%C) for UK papers with 4 authors in eight major fields; averages for the five years 1981, 1986, 1991, 1996 and 2001. (Based on SPSS analysis.)

* This was to keep the numbers of papers below the limit of 65,536 rows in MS Excel.

Figure 3 shows raw data (i.e., without regression analysis) for W/t and %C as a function of numbers of authors per paper, A, for engineering and technology. The trends are similar for all the other fields except for mathematics, for which there are insufficient data for the results to be statistically significant. The use of colons in the titles is higher for single-authored papers (13%), declines to a minimum (8%) for 3-10 authors, and then rises steeply for papers with large numbers of authors. The average number of words per title, on the other hand, rises steadily with the number of authors but soon reaches an apparent plateau (about 11 words) with 4 or more authors. Table 4 shows the percentage use of colons in the titles for different fields for papers with 1, 4 and 8 authors as well as the average numbers of words per titles for 1 and for 4 authors; these data come from the SPSS analysis.

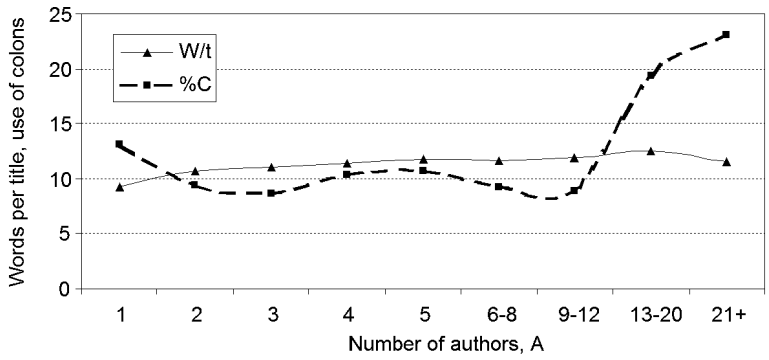

Figure 3. Variation of numbers of words per title (W/t) and percentage use of colons (%C) with number of authors (A) for UK engineering and technology papers; averages for the five years 1981, 1986, 1991, 1996 and 2001. (Raw data.)

Table 4. Variation of colon use (%C) and the number of words per title (W/t) for UK papers with different numbers of authors (A) in seven major fields; averages for the five years 1981, 1986, 1991, 1996 and 2001

Field	%C: A = 1	%C: A = 4	%C: A = 8	W/t: A = 1	W/t: A = 4
Biology	14.1	8.0	11.6	12.0	14.5
Biomedical research	22.9	13.9	13.2	9.9	13.8
Chemistry	25.8	21.3	27.4	9.4	14.2
Clinical medicine	34.5	23.2	23.2	8.7	12.5
Earth and space	16.7	13.6	15.9	10.0	11.8
Engineering & techn.	12.7	6.3	7.6	8.8	10.9
Physics	15.3	11.3	10.0	9.4	11.4

Table 5 shows that there has been an increase in the use of colons in the titles in all the different fields. This increase was fairly steady except for physics, where there was a sudden jump from 9% in 1996 to 25% in 2001. This is because *Physical Review* (sections *A* to *E*, and *Letters*) now lists the number of the article as an integral part of the title, thus: *Search for Dilepton Signatures from Minimal Low Energy Supergravity in PPover Bar Collisions at Root S=1.8 TeV: Art. No. 091102.*

However, the trend in words per title is not uniform, and the overall steady increase observed (from 11.7 to 13.1 words) is occasioned mainly by the inexorable increase in authorship shown in Table 1. Table 5 shows the changes over 20 years (from 1981 to 2001) in the lengths of titles and the use of colons for the individual fields.

Table 5. Changes over 20 years (from 1981 to 2001) in use of colons (%C) and words per title (W/t) for UK papers in eight major fields

Field	Δ %C	Δ W/t
Biology	12.2	–0.3
Biomedical research	3.6	4.0
Chemistry	4.2	–9.5
Clinical medicine	4.2	8.7
Earth and space	11.3	5.2
Engineering & technology	3.1	5.4
Mathematics	1.2	2.0
Physics	13.8	–2.4

Our SPSS analysis of the UK papers in two fields, clinical medicine and biomedical research, by the research level of the journal shows that papers in basic journals have longer titles but fewer colons, compared with ones in clinical journals. The differences are 6.1 words in clinical medicine and 3.6 words in biomedical research, and 23% and 10% in the use of colons, respectively.

Results of Study 2: the world cancer papers

As with Study 1, there is a steady increase in the number of authors per paper over time (see Table 1). And, in parallel with this increase, there is a small but steady increase in the use of colons in the titles, from 23% to 26%, and in the average number of words per title, from 12.6 to 15.0. Figure 4 shows that the use of colons (34% in 2001) by single authors (an increasingly rare breed in cancer research, accounting for only 14% of papers in that year) is greater than that (about 25%) in the papers by 3-10 authors. However, this is then eclipsed by papers with long authorship lists, most of which report large-scale clinical trials and major genetic studies. Figure 4 also shows

that the number of words per title has increased with authorship numbers in much the same way as that shown in Figure 3, except for the 1981 papers. (Here, however, there were only 13 papers with many authors so this particular result is unrepresentative.)

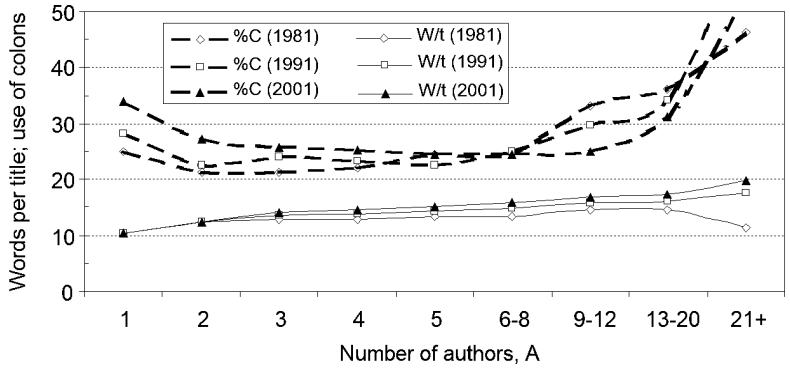

Figure 4. Variation in number of words per title (W/t) and percentage use of colons (%C) with the number of authors (A) in world cancer papers in three years: 1981, 1991 and 2001. (Raw data.)

When considering the research level of the journal and their individual titles we found that the presence of a clinical word in the paper title added an average of 1.4 words to the title length and increased the chance of a colon being present by 4%. In contrast, the presence of a basic word added an average of 2.6 words to the title length but reduced the chance of a colon by 9%. Figure 5 shows that papers in clinical journals (RL = 1.0) had, on average, four less words per title than did papers in basic journals (RL = 4.0). However, journal RL has no significant effect on the use of colons (not shown).

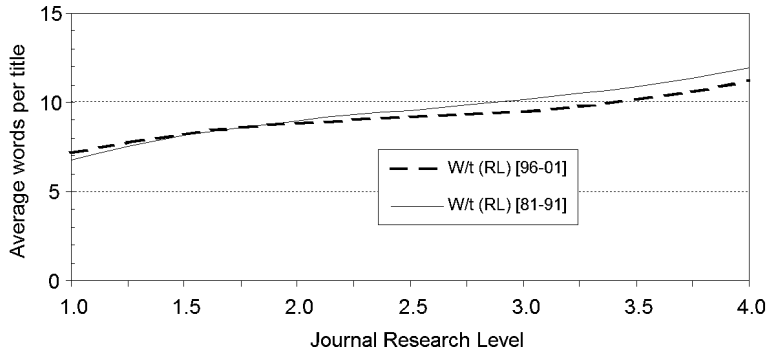

Figure 5. Variation in the numbers of words per title (W/t) as a function of journal RL (1 = clinical, 4 = basic) for world cancer papers in 1981-91 and 1996-2001. (Results calculated from SPSS analysis.)

The results of the SPSS analysis on the effects of author numbers on W/t and %C are similar to those shown in Figure 4 for the raw data. There is a steady rise in W/t with A with an apparent flattening off as A approaches 10 but in fact it continues to rise for papers with very large numbers of authors. The likelihood of colon use is a maximum (48%) for A = 1; it then steadily declines (to 28%) at A = 9 and then rises slowly as A increases still further.

After allowance is made for the increase in author numbers and other parameter changes, the effect of the passage of 20 years on world cancer papers has been to increase the likelihood of colon use by about 1.3% and to increase the number of words per title by 1.1. (The coefficients from the multiple regression analysis of %C and W/t with T obtained from the two sets of papers are almost the same: 0.57 and 0.70 for %C and 0.055 and 0.051 for W/t.) These are both rather small quantities, and show that the observed increase in authorship numbers (see Table 1) has had a somewhat bigger effect. Since almost all cancer research comprises biomedical research or clinical medicine, and both these fields showed an increase in words/title with time (see Table 5), it is to be expected that cancer research would also show this effect.

Finally, as Table 6 shows, the nationality of the authors had little effect on either the percentage of titles with colons or on title length, although there were a few exceptions. This table shows the results from only those countries whose outputs had a significantly different %C or W/t in both the periods analysed (1981–1991 and 1996–2001). Only Italy and Japan appear in both lists.

Table 6. Countries whose papers in cancer research had a significantly different use of colons (%C) or words per title (W/t) in both (1) 1981-91 and (2) 1996-2001

Country	Δ %C (1)	Δ %C (2)	Mean	Country	Δ W/t (1)	Δ W/t (2)	Mean
France	+9.5%	+5.4%	+7.5%	Belgium	+0.6	+1.5	+1.0
Italy	+5.2%	+5.6%	+5.4%	Canada	+0.3	+0.4	+0.3
Japan	– 4.1%	– 4.8%	– 4.5%	Denmark	+0.7	+1.1	+0.9
				Germany	+0.5	+0.3	+0.4
				Italy	– 0.5	– 0.2	– 0.4
				Japan	+0.6	+0.5	+0.6
				Netherlands	+1.1	+0.7	+0.9
				Sweden	+0.4	+0.3	+0.3

Discussion

The main findings of this paper are common to both Studies 1 and 2. It appears that:

1. Titles have been increasing in length over time.
2. The average number of authors per paper has been increasing over time.

3. The percentage of titles with colons has been increasing over time but the amount varies within different fields.
4. Single authors produce more titles with colons than do multiple authors (in nearly all fields), but this reverses when the number of co-authors is high.
5. Within clinical medicine and biomedical research, papers in basic research journals tend to have longer titles, but use fewer colons, than ones in clinical journals.

In addition, the results of Study 2 show that these results are largely unaffected by the nationalities of the authors concerned.

Findings 1 and 2 replicate and extend previous findings on title length and the numbers of authors involved as discussed in the Introduction. Finding 3 documents further disciplinary differences (within the sciences) and the growth of titles with colons through time. Finding 4 is perhaps the most surprising in this research and merits further comment.

As noted earlier the publications with large numbers of authors in the cancer papers come from large-scale clinical trials and major genetic studies. In the science publications large numbers of authors are currently associated primarily with nuclear and particle physics (with papers from CERN in Switzerland sometimes having many hundred authors), and to a less extent with clinical medicine, biomedical research and earth & space. Just why these kinds of work (as compared to others) should employ titles with colons is a bit of a mystery, but one might speculate that the format allows the title to appear more embracing. Here are some typical examples from physics: *Developments for Radiation Hard Silicon Detectors by Defect Engineering: Results by the CERN Rd48 Rose Collaboration; Quadrupole Moments of High Spin Isomers: Test of the Tilted Axis Cranking Model.*

Although in both Studies 1 and 2 single authors use more colons in their titles than do moderate numbers of co-authors, this result should not be overstated. The difference, after all, is only one of about 5% for the UK science papers and about 10% for the cancer papers. Further, it has to be remembered that not all single-authors consistently use titles with colons and, indeed, that they might vary within themselves. Table 7 shows that both the authors of this present paper have published single- and co-authored papers, and that they have used different formats in both situations.

Table 7. Types of titles used by the present authors when writing alone and with co-authors

Title type	L alone	L + co-auth.	H alone	H+ co-auth.
Simple sentence	16	21	8	3
Sentence with colon	6	6	2	7
Two sentences	0	1	4	3

HARTLEY (2005) attempted to describe (and count) papers in psychology in terms of the balance between the two parts in titles with colons. Thus he discriminated between what he called:

'short: long': e.g., *Notion of a document: a center of gravity attraction for getting metricians together;*
'long: short': e.g., *Patents, research and development expenditures, country size, and per capita income: an international comparison,* and
'balanced' e.g.,. *A bibliometric profile of top scientists: a case study in chemical engineering.*

For the 336 psychology papers with titles with colons, Hartley found 'short:long' titles in 45%, 'long:short' titles in 32%, and 'balanced' titles in 23% of the rest. This parallels our current findings for *Scientometrics* (1994-2003). Here we have a similar balance of 'short:long' titles (53%), long:short titles (44%) and 'balanced' titles (9%). However, when we analysed the UK papers from different fields in 1996, we found that the balance between short:long and long:short titles varied with the field, see Table 8.

Table 8. Balance of two-part titles between length of first part and length of second part for UK papers in different fields, 1996

Field	Total	Short:long, %	Long:short, %	Balanced, %
Physics	574	35.0	57.0	8.0
Earth & space	702	35.9	56.6	7.5
Biology	579	40.9	51.1	7.9
Biomedical res	1587	45.9	47.1	7.0
Clinical med	4726	45.4	46.5	8.1
Engr & techn	314	46.8	44.9	8.3
Chemistry	1222	49.4	43.8	6.8
Other	265	53.2	38.9	7.9
Mathematics	47	55.3	31.9	12.8
Total	10016	44.8	47.5	7.7

Other researchers have used a more linguistic approach to classifying titles with colons. Thus, for example, SWALES & FEAK (1994) distinguished between four types of titles with colons according to how they separated ideas: these were: problem:solution; general:specific; topic:method; and major:minor. However, they did not report any data on the relative proportions of these formats. To do this for the present study would currently require hand-counting and is beyond anyone's capabilities with the present data.

It would be useful to ask authors about their practices in choosing titles when writing papers singly or with others. It may be that, as single authors in their papers acknowledge the help of others more often than do multiple authors (HARTLEY, 2003), single authors might find themselves explaining to others what their paper is about in terms of first general and then specific issues, whereas co-authors already know this and

do not have to explain it to each other. But this is speculation. It would also be interesting to know more about how journal editors, referees and readers respond to titles in different formats, and why some formats (involving the use of questions, for example) are largely eschewed in scientific papers. HYLAND (2002) only found them occurring in the 'soft' as opposed to the 'hard' disciplines. ANTHONY (2001) only found two titles in the form of questions in a sample of 600 articles in computer science.

A particularly interesting aspect of this paper is that it has again been possible to employ the actual words used in the titles of the articles in the cancer journals to classify the nature of the article and the journal in which these articles appear. Thus we have been able to compare the use of colons in basic and clinical papers. We originally envisioned that we might be able to rank the science journals in Study 1 along similar lines – from basic, such as maths and physics, through to applied, such as engineering and technology. We expected, on the basis of our reading of the literature outlined in the Introduction, that the titles would be longer and that the authors would use more colons in the applied articles. However, this did not fully materialise. For the cancer papers, contrary to our predictions, the basic papers have longer titles than the applied ones (v.s., Figure 5), and the data on colon use are inconclusive. Within the UK, biology has the longest titles and mathematics the shortest; both are rather basic disciplines. Colon use is highest in clinical medicine, a relatively applied field, and lowest in engineering and technology, also an applied field. Of course simply ranging disciplines along a continuum from 'basic to applied' has its own problems, as there are ranges within this dimension for each discipline (HARTLEY et al., 2004). Thus, for example, we have 'pure' and 'applied' maths, and psychology embraces both neuropsychology and counselling. It may be that it is these variations within the disciplines that led to our mixed results.

Finally, we note with interest our finding from Study 2 – that the results do not seem to be greatly affected by the nationalities of the authors concerned. Table 6 shows the pertinent results, where it can be seen that only 8 of the 20 countries (listed in Table 3) differed significantly from the mean in both periods in terms of title length, all except Italy having longer titles than average, and only three differed in their use of colons. These results suggest that research in the sciences is now so international in its language that national differences in the titles of articles are no longer noticeable. [Almost all the titles were originally written in English.] There was, however, a small group of northern European countries that tended consistently to have slightly longer titles (Sweden, Denmark, Germany, Belgium and the Netherlands); it would be interesting to know if there are cultural factors underlying this usage.

*

We are grateful for helpful comments by Dr. Helmut Abt.

References

ANTHONY, L. (2001). Characteristic features of research article titles in computer science. *I.E.E.E. Transactions on Professional Communication,* 44 (3) : 187–194.

BERKENKOTTER, C., HUCKIN, T. N. (Eds) (1995). *Genre Knowledge in Disciplinary Communication.* Hillsdale, N.J.: Erlbaum.

CRONIN, B. (2001). Hyperauthorship: A postmodern perversion or evidence of a structural shift in scholarly communication practices? *Journal of the American Society for Information Science and Technology,* 52 (7) : 558–569.

DAY, R. A. (1998). *How to Write and Publish a Scientific Paper* (5th edition). Westport, Connecticut: Greenwood.

DIERS, D., DOWNS, F. S. (1994). Colonizing: a measurement of the development of a profession. *Nursing Research*, 43 (5) : 316–318.

DILLON, J. T. (1981). The emergence of the colon: An empirical correlate of scholarship. *American Psychologist,* 36 (8) : 879–884.

DILLON, J. T. (1982). Impact of the colon: Preliminary reactions. *American Psychologist,* 37 (6) : 716.

FONTANET, I., COLL, J. F., PALMER, J. C., POSTEGUILLO, S. (1997). The writing of titles in academic research articles. In: R. M. CHAMORRO, A. R. NAVETTE (Eds), *Lenguas Aplicadas a las Ciencias y la Tecnologia Approximaciones.* Caceres, Spain: Universidad de Extremadura, Servicio de Publicaciones, pp. 155–158.(Cited by ANTHONY, 2001).

FONTANET, I., POSTEQUILLO, S., COLL, J. F., PALMER, J.C. (1998). Linguistic analysis of research article titles: Disciplinary variations. In: I. VAZQUEZ, I. CAMILLEU (Eds), *Perspectivas Praguietices en Linguistics Applicada.* Zaragozs, Spain: Anubar. pp. 443–447. (Cited by ANTHONY, 2001).

HALL, G. M. (Ed.) (2003). *How to Write a Paper* (3rd edition). London: British Medical Journal Publishing Group.

HARTLEY, J. (2003). Single authors are not alone: Colleagues often help. *Journal of Scholarly Communication,* 34 (2) : 108–113.

HARTLEY, J. (2005). To attract or to inform: What are titles for? *Journal of Technical Writing and Communication,* 35.

HARTLEY, J., SOTTO, E., FOX, C. (2005). Clarity across the disciplines: An analysis of texts in the sciences, social sciences and arts and humanities. *Science Communication,* 26 (2) : 188–210.

HYLAND, K. (2002). What do they mean? Questions in academic writing. *Text,* 22 (4) : 529–557.

KILGOUR, F. G. (2004). An experiment using coordinate title word searches. *Journal of the American Society for Information Science and Technology,* 55 : 74–80.

KOSTOFF, R. N., BLOCK, J. A., STUMP, J. A., PFEIL, K. M. (2004). Information content in Medline record fields. *International Journal of Medical Informatics,* 73 (6) : 515–527.

KUTCH, T. D. C. (1978). Relation of title length to numbers of authors in journal articles. *Journal of the American Society of Information Science,* 19 (4) : 200–202.

LEWISON, G. (1999). The definition and calibration of biomedical subfields. *Scientometrics,* 46 (3) : 529–537.

LEWISON, G., PARAJE, G. (2004). The classification of biomedical journals by research level. *Scientometrics,* 60 (2) : 145–157.

LIU, Z. (2002). Trends in transforming scholarly communication and their implications. *Information Processing and Management,* 39 (6) : 889–898.

MABE, M. A., AMIN, M. (2002). Dr. Jekyll and Dr. Hyde: Author-reader asymmetries in scholarly publishing. *Aslib Proceedings,* 54 (3) : 149–157.

MEADOWS, J. (1998). *Communicating Research*. San Diego: Academic Press.

MICHELSON, G. (1994). Use of colons in titles and journal status in industrial relations journals. *Psychological Reports,* 74 : 657–658.

O'CONNOR, M. (1991). *Writing Successfully in Science.* London: HarperCollins.

PERRY, J. A. (1985). The Dillon hypothesis of titular colonicity. *Journal of the American Society for Information Science,* 36 (4) : 251–258.

SWALES, J. M., FEAK, C. B. (1994). *Academic Writing for Graduate Students: A Course for Nonnative Speakers of English.* Ann Arbor, Michigan: University of Michigan Press.

TOWNSEND, M. A. R. (1983). Titular colonicity and scholarship. *New Zealand Journal of Psychology*, 12 : 41–43.

WHISSELL, C. (2004). Titles of articles published in the journal Psychological Reports: Changes in language, emotion, and imagery over time. *Psychological Reports*, 94 : 807–813.

YITZHAKI, M. (1994). Relation of title length of journal articles to number of authors. *Scientometrics,* 30 (1) : 321–332.

YITZHAKI, M. (2002). Relation of the title length of a journal article to the length of the article. *Scientometrics,* 54 (3) : 435–477.