

This item was submitted to Loughborough's Institutional Repository (<https://dspace.lboro.ac.uk/>) by the author and is made available under the following Creative Commons Licence conditions.



**CC creative commons**  
COMMONS DEED

**Attribution-NonCommercial-NoDerivs 2.5**

**You are free:**

- to copy, distribute, display, and perform the work

**Under the following conditions:**

**BY:** **Attribution.** You must attribute the work in the manner specified by the author or licensor.

**Noncommercial.** You may not use this work for commercial purposes.

**No Derivative Works.** You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

**Your fair use and other rights are in no way affected by the above.**

This is a human-readable summary of the [Legal Code \(the full license\)](#).

[Disclaimer](#) 

For the full text of this licence, please go to:  
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

# What do UK academics cite? An analysis of references cited in UK scholarly outputs

Claire Creaser and Mark AC Summers<sup>1</sup>

*LISU, Loughborough University, Loughborough, UK, LE11 3TU*

Charles Oppenheim

*Department of Information Science, Loughborough University, Loughborough, UK, LE11 3TU*

**Abstract** This study used a bibliometric method to find quantitative evidence of publication and citing patterns within UK academia. The publications of a random sample of UK research-active academics for each of the years 2003 and 2008 were collected and analysed to gather data regarding referencing practices, along with any identifiable trends between the two years. References were categorised by type of material to show the proportions of each type used. Comparisons between the two years showed that the use of journal articles had increased. There was also an increase in the average number of publications per author. A large number of authors had no publications in the target years. (109 words)

**Keywords:** *Publication behaviour; Research assessment; Disciplinary differences; Cited references*

---

<sup>1</sup> Correspondence to: Mark Summers, LISU, Loughborough University, Loughborough, UK, LE11 3TU, [m.a.c.summers@lboro.ac.uk](mailto:m.a.c.summers@lboro.ac.uk).

## Introduction

This paper reports on one aspect of a project which investigated the publication and dissemination behaviour of UK researchers and the effect of research assessment on this behaviour<sup>2</sup>. Using a bibliometric method to provide quantitative evidence regarding publication and citing patterns within UK academia, we investigated:

- The types of output and means of dissemination used by researchers in different broad disciplines, and how this changed between 2003 and 2008;
- The number and type of items that authors cited in their work, how this varied by discipline, and whether it had changed over time.

There is a generalisation in the study of research communication, that scientists make the most use of journals, and humanities scholars, the most use of books. This is, to some extent, an oversimplification, and a varied picture of the use of different material has been shown in previous studies of publication practices. In general, these have concentrated on specific subject areas and/or countries for their analyses and used a variety of methods and data sources.

Books and book chapters have been shown to account for varying proportions of cited literature in a number of subjects in the social sciences and humanities, accounting for 74% of references in literature research (Nederhof 1995) and 85% in philosophy (Cullars 1998). Lindholm-Romantschuk & Warner (1996) and Broadus (1971) found differing proportions of references to books in Sociology of 72% and 62% respectively. Cronin et al. (1997) also provided qualitative evidence of strong use of books in Sociology. However, a number of studies have found that journals are the favoured means of publication in many disciplines in the social sciences and humanities (including Burnhill & Tubby-Hille 1994, Hargens 2000 and Nederhof 2006).

Journal use was reported to be high in the sciences, for instance, shown by Glänzel & Schoepflin (1999) to be 85% of references in solid state physics and 84% in analytical chemistry. This is supported by an analysis of the ISI Science Citation Index by Moed (2005), who worked “under the assumption that references in ISI source journals adequately reflect the importance of written communication media”, and found that there was excellent coverage of the ‘hard’ sciences. This was also found to be the case for the Scopus index (Lancho-Barrantes et al. 2010).

Publication behaviours are not fixed and have been seen to change. Kyvik (2003) shows an increase in journal use in the social sciences and humanities; however a recent study by Larivière et al. (2006) suggested in contrast that journal use was “stagnant and has even tended to diminish slightly in the 1990s”. Nederhof (2006) concluded that there exists “a tendency that science communication patterns are increasingly found in social sciences and humanities, [although] differences in publication and citation behaviour still tend to be large in many fields”.

## Method

This study focused on the citing behaviour of UK academics, and used a novel method to ascertain the numbers and types of material cited. The raw data comprised lists of published works, with an indication of what type of material those works were. As it was not the intention to investigate links between specific works, the data collected took no account of the authorship of cited works. As far as the authors are aware, there are no previous studies that use this method.

In outline, the method adopted for this study was to:

- Obtain a first sampling frame of UK academics with outputs in 2003;
- Obtain a second sampling frame of UK academics with outputs in 2008
- Draw a stratified random sample from each list;
- Obtain details of the published outputs for that year for each member of the sample;
- Obtain, examine and categorise the items in the bibliography of each output.

---

<sup>2</sup> Communicating knowledge: how and why UK researchers publish and disseminate their findings  
<http://www.rin.ac.uk/our-work/communicating-and-disseminating-research/communicating-knowledge-how-and-why-researchers-pu>

## Sampling frame and process

The sampling frames used were the submissions to the UK Research Assessment Exercise (RAE)<sup>3</sup> for 2001 and 2008. These data are collected from UK higher education (HE) institutions, by discipline (Unit of Assessment – UoA<sup>4</sup>), in order to inform the allocation of research monies to those institutions. Whilst academics submitted in the RAE had been judged to be research active at the relevant time for this study (generally having at least four high quality research outputs over the period since the previous assessment), there was no guarantee that authors selected from these lists would necessarily have outputs in the specific year of interest: 2003 or 2008.

The objective was to obtain two independent samples of 400 authors, one for each year of publication. This sample size was chosen based on standard statistical theory (Thompson 2002), balanced against the need to complete the work within the project timescale. A sampling design stratified by broad discipline was used, with the aim of producing a sufficiently large sample in the smaller discipline groups for comparative analysis (Table 1). The sampling frame was sorted by UoA into broad discipline groups, and a systematic sample (statistically equivalent to a simple random sample) drawn from each. A supplement of 30% over the desired sample size was drawn, to allow for authors who might have no outputs in the relevant year. Details of the RAE Units of Assessment and Higher Education Statistics Agency (HESA) academic cost centres allocated to each broad discipline can be seen in Fry et al (2009a).

**Table 1** Distribution of UK academics by broad subject discipline

	HESA <sup>a</sup> distribution	RAE 2001 distribution	RAE 2008 distribution	Average of these	Desired sample size
Bio-medicine	31%	24%	21%	25%	75
Physical sciences	10%	15%	15%	13%	55
Engineering	13%	12%	12%	12%	55
Social sciences	21%	22%	26%	23%	75
Humanities	10%	16%	16%	14%	60
Arts	8%	6%	6%	7%	40
Education	9%	5%	4%	6%	40
	100%	100%	100%	100%	400

Notes:

Columns may total more than 100% due to rounding

<sup>a</sup> Based on FTE academic staff 2006/07

Academic staff who changed institution during the assessment period prior to the 2001 and 2008 RAE returns were eligible to be included by both institutions. In order to maintain the equivalence to simple random sampling within each discipline, suspected duplicate records (based on name and UoA) were removed from the author lists prior to sampling.

In practice, the 30% supplement proved insufficient to provide a final sample of 400 authors with outputs in each relevant year, but the project had insufficient time to extend the samples. This suggests that, if RAE lists or their equivalent are used to sample academic staff for this type of analysis, a larger initial sample would be required to allow for individuals with no outputs in the year of interest. Subsequent analysis indicated that, owing to the relatively small number of text-based research outputs found, an additional bias towards the Arts would also be desirable, in order to allow this area to be included in valid comparisons of practices across disciplines (this is discussed further below).

<sup>3</sup> The UK higher education funding bodies have held six research assessment exercises since 1986 in order to distribute research-specific funding on the basis of research quality.

<sup>4</sup> At the last exercise in 2008, each subject studied within UK higher education was assigned to one of 67 Units of Assessment which in turn were assigned to one of fifteen main panels that group “broadly cognate disciplines whose subjects have similar approaches to research” (HEFCE et al. [n.d.]a)

## Data collection

Searches were undertaken to obtain lists of outputs from the sampled authors in each year. All forms of dissemination were included, not only those which contribute to assessment. For practical reasons, searching was split into two sections: those subjects with good coverage in ISI Web of Science (WoS) (Bio-medicine, science and engineering) and those with less or minimal coverage (social science, humanities, arts and education). WoS searches were performed using the selected author's name and target year, limited to those with a UK address. For academics in subject areas that were not well covered by WoS, the data collection was carried out manually, using a variety of sources, including researcher/departmental websites, bibliographic databases, Google Scholar, the RAE submissions database and the wider internet.

For WoS-sourced outputs, when relevant entries were identified in the search results, the indexed metadata and cited reference lists were printed and references manually categorised and counted. For the non-WoS subjects, a list of outputs identified as being produced by each author was drawn up, copies of the items were sourced, either in electronic or physical format, and the references extracted.

All cited items were categorised as follows:

- journal articles
- conference papers
- theses
- grey literature (e.g. technical reports; working papers; occasional papers; Governmental/NGO publications; British & international standards; mimeos)
- books (including book chapters)
- websites
- items in press or forthcoming
- other items (e.g. non textual material; data sets; Parliamentary statutes; patents; historical documents; archive material)

Items described as 'in press' or 'forthcoming' were kept as a separate category, as it was felt that these cannot be relied upon as existing as publications, as they may not eventually appear.

Table 2 summarises the achieved sample in terms of the numbers of authors included in the analysis, the numbers of those for whom outputs were found in each relevant year, the total number of outputs analysed, and the number of references counted in those outputs.

**Table 2** Summary of data collected

	2003				2008			
	Authors		Outputs		Authors		Outputs	
	Sample size	No with outputs	No. outputs analysed	No. references identified	Sample size	No with outputs	No. outputs analysed	No. references identified
Bio-medicine	97	45	102	2535	96	52	235	7117
Physical sciences	72	49	223	5794	70	49	212	7001
Engineering	72	27	87	1347	70	44	149	3311
Social sciences	98	37	75	4799	96	52	137	5988
Humanities	78	54	81	5575	77	40	34	1315
Arts	52	14	12	416	51	21	17	100
Education	52	26	51	1938	51	22	37	1262
	521	252	631	22404	511	280	821	26094

## Challenges

A number of challenges emerged during the data collection stages that affected the analysis. The most significant of these are described below.

### *Identifying the correct researcher*

Researchers were identified from their surname, initials, institutional affiliation and UoA. In the majority of cases this was sufficient, although there were occasional difficulties experienced in

locating the correct individual. In particular, the 2003 sample was drawn from the RAE2001 list that included all researchers returned from all categories. Thus a small number of names with little or no continuing connection to UK higher education were included in the sample.

### *Compiling accurate reference lists*

There was no practical way of assessing the completeness of the lists of outputs obtained. Even where a researcher provided a detailed list of research outputs on their webpage, such lists could not be relied upon to be complete or up to date. Lists often stopped at some point in the recent past, notably the end of 2007 (the end of the RAE2008 census period). Furthermore, adequate detail was not universal, with the majority of researchers sampled listing only selected outputs, or none at all, on their web pages. Furthermore, some references on these lists were found to be incorrect. The most common errors concerned publications that were, presumably, forthcoming at the time of compiling the list but which were eventually published in a different year.

### *Researchers with no analysable output in the appropriate year*

Despite the universally acknowledged pressure to publish, a surprisingly large number of the sampled researchers showed no identifiable output in the appropriate year. This was the case even though an inclusive definition of output was employed, following the lead of the RAE rules. In particular, the Arts showed a notable lack of visible output. Here, although 22 research outputs were identified from 14 authors in 2003, and 46 from 21 authors in 2008, the number analysed is rather smaller, either because they were not text-based (for example, artistic works, music compositions, performances) or because the research team were unable to obtain a copy of the work for examination.

In both years, especially for those names drawn from the Art and Design UoA, most appeared to be practising artists<sup>5</sup>, who had minimal presence on departmental websites, with the wider internet being only a little help. Without doubt, in this case many, if not most, of the artists and musicians will have been producing work that was accessible by the wider world but this is not necessarily text-based, and did not show up in the search results.

Table 3 gives details of the number and proportion of researchers for whom no outputs could be identified in 2003 and 2008.

Table 3. Authors with no discoverable publications in the relevant years

	2003		2008	
	Number	% of sample	Number	% of sample
Bio-medicine	52	54%	44	46%
Physical sciences	23	32%	21	30%
Engineering	45	63%	26	37%
Social sciences	60	61%	44	46%
Humanities	24	30%	37	48%
Arts	38	73%	30	59%
Education	26	50%	29	57%
Total	269	52%	231	45%

### *Output without citation*

A number of items examined contained no references, for example book reviews, conference outputs that were not published in formal proceedings or, in the case of music, CDs. This was more true of the Arts and Humanities than other disciplines, possibly because of a more informal approach to research gatherings, whether conferences, symposia etc.

<sup>5</sup> Submissions to UoA 63 (Art and design) show the majority of the sampled authors to be producing artifacts/exhibitions rather than written outputs. (HEFCE et al. [n.d.]b).

### *Inability to obtain a copy of outputs*

When attempting to retrieve non-WoS material, occasionally it proved impossible to obtain a copy of an item for examination within the resources of the project

### *Reference counting*

WoS cited references came in two forms, those indexed by WoS and those not indexed. The former were straightforward and identification of item type cited was simple, whilst the latter required more skill. WoS truncation of non-indexed titles can often hide the identity of the cited work, but in the main an internet search provided enough information for the identification of problematic references.

In non-WoS reference lists, the categorisation of type was more subjective. Case by case decisions included whether an item was a book or grey literature; a book or historical pamphlet; grey literature or website; journal article or book (when one was republished as another); etc.

Conference outputs were identified where possible, in whatever form they occurred. In particular, identification of websites was problematic, given that much material was accessed via the WWW but was not a website per se, but another type of output in electronic format.

### **Analysis**

As publication and citation practice varies between broad disciplines, any difference in the distribution across disciplines between the samples for the two years may influence the outcomes of the longitudinal analysis. Similarly, in order to draw inferences about the population of UK academics as a whole, the effect of disciplinary differences between the sample and the population must be removed. All analyses were therefore carried out on weighted data. Weights were calculated based on the percentage of authors analysed divided by the average percentage of academic staff in each discipline<sup>6</sup>. The weighted data therefore reflect the overall distribution of UK academics by discipline. Table 4 details the numbers of authors whose research outputs were examined and included in the analysis, and shows the weights applied to each discipline in the analysis. The total number of outputs analysed across both years was 1,452.

Table 4. Calculation of weights

	<i>Proportion of population (from table 1)</i>	2003			2008		
		Authors Analysed	Proportion of sample	Weight	Authors analysed	Proportion of sample	Weight
Bio-medicine	25%	44	19%	1.36	52	21%	1.22
Physical sciences	13%	52	22%	0.60	47	19%	0.70
Engineering	12%	27	11%	1.07	45	18%	0.68
Social sciences	23%	37	16%	1.45	51	20%	1.12
Humanities	14%	42	18%	0.77	20	8%	1.71
Arts	7%	9	4%	1.74	15	6%	1.11
Education	6%	24	10%	0.57	19	8%	0.76
	100%	235	100%		249	100%	

The analysis was carried out using the SPSS® software package. Formal tests were carried out for statistical significance of apparent differences between years, between disciplines (excluding Arts), and, where there were sufficient data for the tests to be valid, between types of source material. All results reported as being statistically significant had a less than 5% probability of occurring by chance; the significance levels are noted in the text.

<sup>6</sup> For example, Bio-medicine for 2003:

Percentage of academic staff = 25.4%    Percentage of sample = 18.7%  
Weighting – 25.4/18.7 = 1.36

# Results

## Types of output

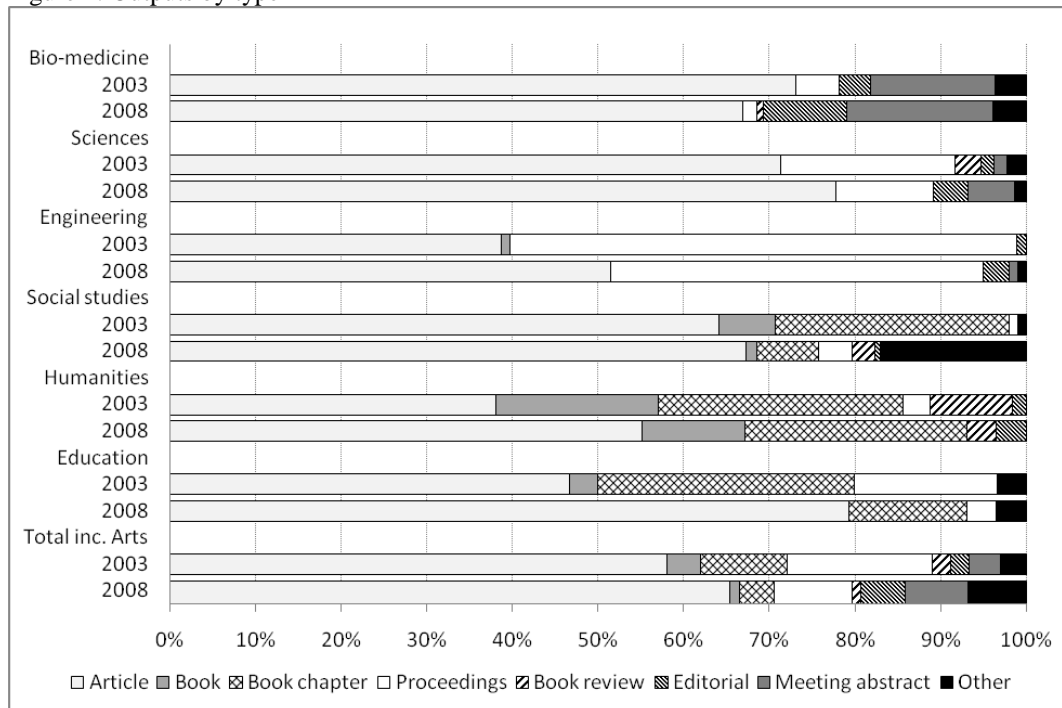
Table 5 shows the average number of outputs, of all types, found per author in each year, by discipline. There was a small, but statistically significant, increase in the average from 2.5 in 2003 to 3.2 in 2008. Much of this is due to a considerable increase in Bio-medicine, which accounts for approximately one quarter of all UK HE researchers. Examination of the detailed distributions shows that this increase is caused by a general extension of the tail of the distribution, rather than by individual outliers.

Table 5. Average number of outputs per author in each discipline

	2003		2008		Significant?
	Mean	Standard error	Mean	Standard error	
Bio-medicine	2.32	.188	4.52	.597	p<0.01
Physical sciences	4.29	.851	4.51	.609	No
Engineering	3.22	.659	3.31	.474	No
Social sciences	2.00	.198	2.69	.266	p<0.05
Humanities	1.93	.303	1.70	.124	No
Education	2.17	.631	1.88	.245	No
Total, inc. Arts	2.50	.170	3.19	.206	p<0.05

Figure 1 illustrates the proportions of outputs by type for each discipline. There were insufficient data to test the apparent differences by discipline, either overall or within year. Combining disciplines, the difference between years overall was found to be statistically significant (p<0.01). There was a greater proportion of journal articles produced in 2008, and of editorial material, meeting abstracts, and 'other' types of material; conversely there were lower proportions of books, book chapters, conference proceedings and book reviews produced. (Note that the data have been weighted to reflect the population distribution of disciplines, so that changes in the disciplinary distribution do not account for the difference over time.)

Figure 1. Outputs by type





## Citation practice

Figures 2 and 3 summarise the data gathered on citation practices. Overall, 6.8% of the textual works examined did not include any references in 2003, compared to 10.8% in 2008. This difference overall was found to be statistically significant ( $p < 0.05$ ). Figure 2 illustrates this by discipline; apparent differences between years for individual disciplines were found not to be statistically significant where there was sufficient data to carry out the test. Although there was insufficient data to formally test the apparent variation, the inclusion or not of references also tended to be associated with the type of output, with meeting abstracts, editorial pieces and book reviews least likely to include references (Figure 3).

Figure 2. Inclusion of references by discipline and year

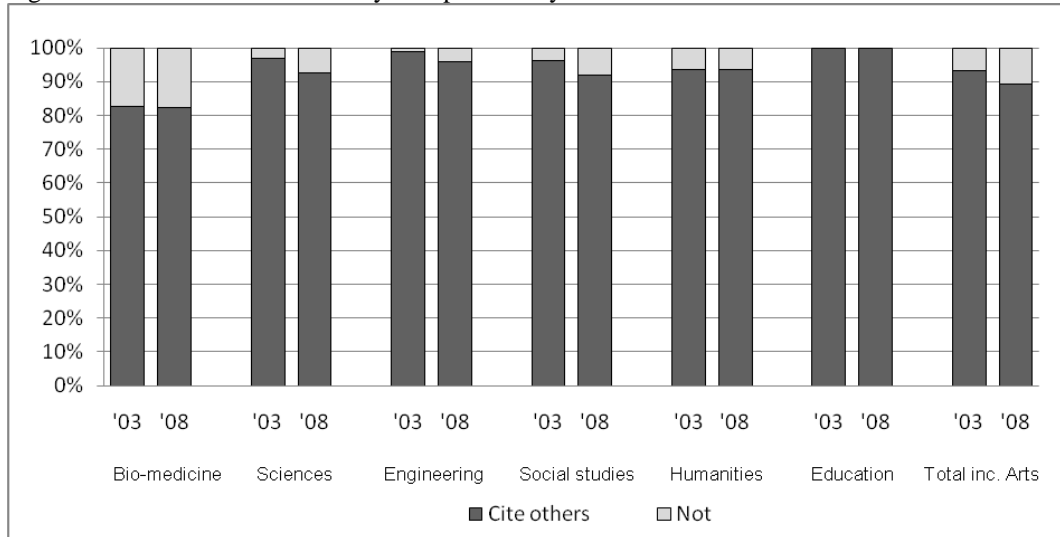


Figure 3. Inclusion of references by source

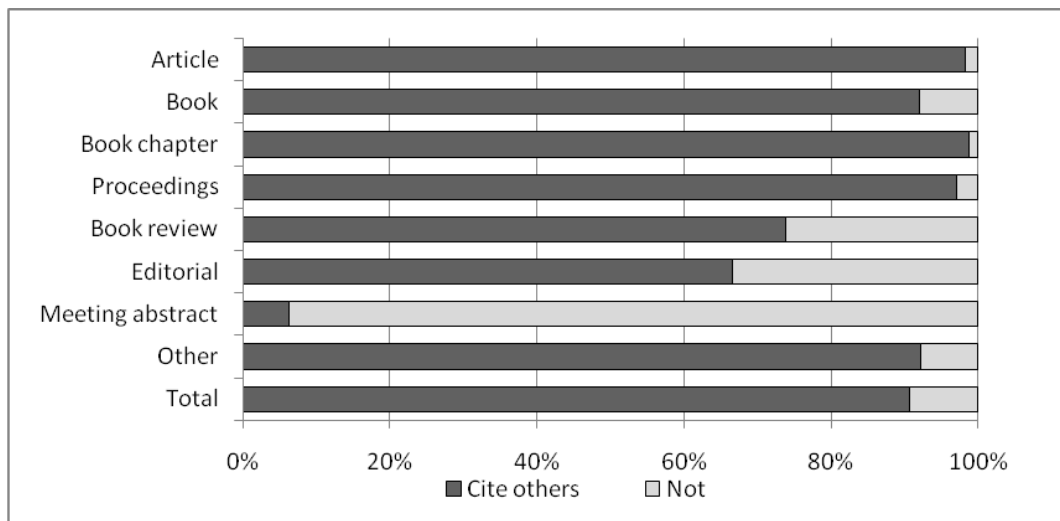


Table 6 shows the average number of references per output, both by discipline and by output type. Overall, the difference in the average number of references per output between 2003 and 2008 was found not to be statistically significant. There are, however, differences in the average number of references per output between types of output and, consequentially, between disciplines.

Table 6. Average number of references per output

	Mean	Standard error	Significant differences?
2003	37.07	2.600	} No
2008	33.68	1.141	
Bio-medicine	28.51	1.259	} p<0.01
Physical sciences	29.68	1.498	
Engineering	18.99	.963	
Social sciences	53.13	3.246	
Humanities	59.75	10.929	
Arts	39.36	14.337	
Education	36.67	5.624	
Article	37.55	.852	} p<0.01
Book	230.96	42.537	
Book chapter	47.16	4.560	
Proceedings	19.48	1.575	
Book review	3.95	1.546	
Editorial	8.90	1.731	
Meeting abstract	1.16	.706	
Other	28.52	5.268	
Overall	35.14	1.293	

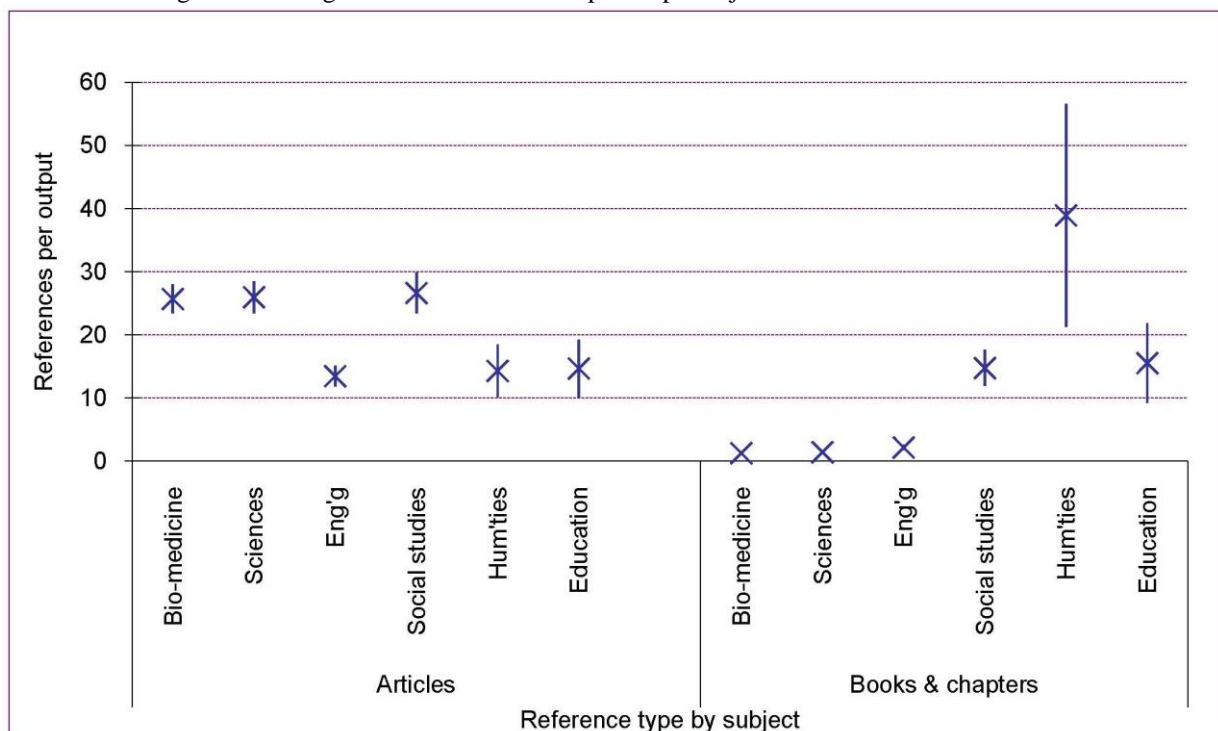
Constructing a model including both source and discipline, analysis of variance indicates that there is also a statistically significant interaction between these factors. This suggests that the average number of references per output for the various types of source outputs differ according to discipline. However, this result should be treated with caution, as there are several discipline/source type combinations that were not observed in the data collected. It seems likely that one potential explanation for the interaction effect observed may be that the data collection method resulted in the identification of almost no books as published research outputs in the Bio-medicine, Science and Engineering disciplines, artificially producing this effect. Although there is no difference over time in the total number of references per output, there are statistically significant differences in the types of work being cited, shown in Table 7. There are more references to journal articles in 2008 than in 2003, but fewer references to books and grey literature. There are also differences in the average numbers of references of different formats between disciplines ( $p < 0.01$  for all types of cited material), and between types of source output ( $p < 0.01$  for all types of cited material except outputs in press, where  $p < 0.05$ ).

Table 7. Average number of references to each type of cited output

	2003		2008		Significant?
	Mean	Std. Error	Mean	Std. Error	
Articles	20.0	0.94	24.3	0.83	p<0.01
Books	11.6	1.89	5.5	0.53	p<0.01
Conference outputs	0.9	0.15	0.8	0.09	No
Grey literature	2.1	0.35	1.2	0.12	p<0.05
Websites	0.3	0.06	0.3	0.10	No
Theses	0.2	0.03	0.2	0.02	No
In press	0.2	0.02	0.1	0.02	No
Other	2.0	0.38	1.2	0.17	p<0.05
Total	37.1	2.60	33.7	1.14	No

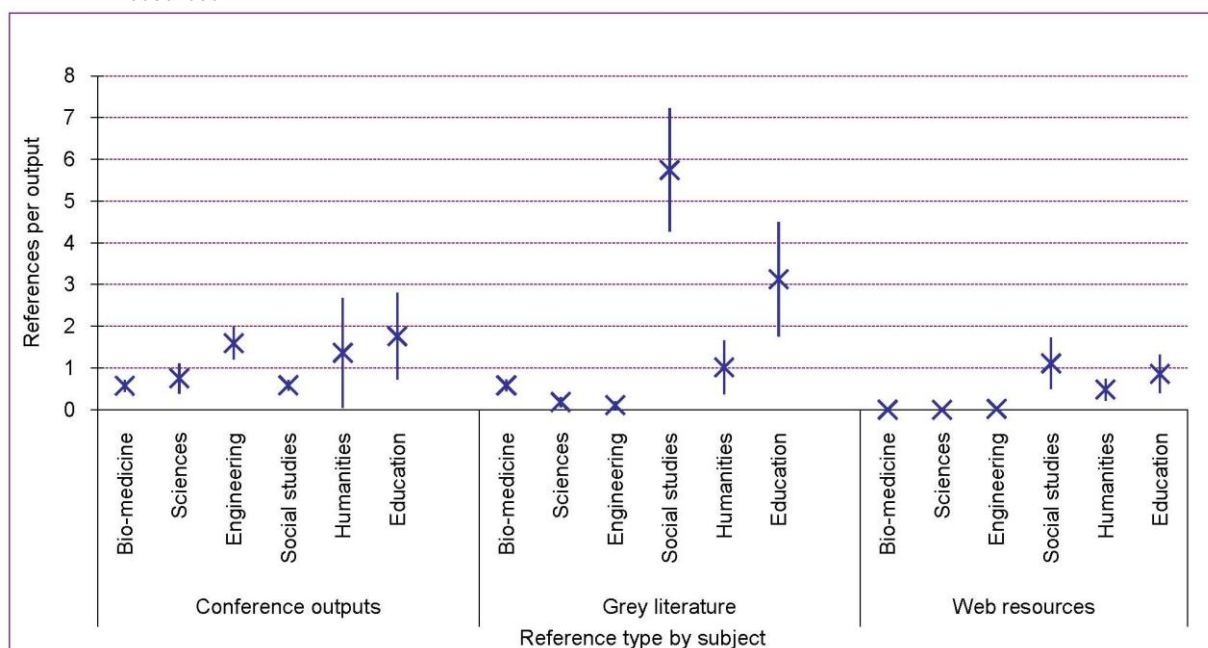
Figures 4 and 5 illustrate the mean number of references per output by discipline, for those types of material cited where the differences between subjects were found to be statistically significant ( $p < 0.01$  in all cases). While all disciplines cite journal articles, Bio-medicine, Physical sciences and Social sciences cite around twice as many, on average, as Engineering, Humanities and Education (Figure 4). Humanities cite the most books, on average, with Social sciences and Education also citing these to some extent (Figure 4). Engineering cites more conference outputs than do Bio-medicine, Physical sciences or Social sciences (Figure 5). Social sciences cite the most grey literature, on average, with Education also citing these, and to a lesser extent, Humanities also (Figure 5). Web sites are rarely cited, but Social sciences, Education and Humanities are more likely to do this than the other disciplines (Figure 5).

Figure 4. Average number of references per output to journal articles and books



X is the mean; the bars indicate the standard error of the mean

Figure 5. Average number of references per output to conference outputs, grey literature and web resources



X is the mean; the bars indicate the standard error of the mean

Table 8 shows the proportions of material of different types cited in a selection of output types. While journal articles are the most frequently cited works overall, books and book chapters are most likely to cite books/book chapters. Conference outputs are most likely to be cited in other conference outputs. There was insufficient data available to investigate the potential effects of disciplinary differences in dissemination behaviour on the patterns observed.

Table 8. Proportion of references, by type of material cited and output type

Citing work	Type of material cited					Total
	Articles	Books*	Conf'ce	Grey lit	Other	
Article	74%	15%	2%	3%	5%	100%
Book	27%	64%	1%	5%	4%	100%
Book chapter	31%	47%	2%	8%	11%	100%
Conference output	71%	10%	8%	2%	9%	100%
Editorial	84%	11%	1%	2%	2%	100%
All outputs	64%	23%	2%	4%	6%	100%

\*Both books and book chapters

## Discussion and conclusions

This bibliometric study was designed to cover the full range of research in the UK, and the data collection was an extensive undertaking, which could not be automated. The time required to undertake such exercises should not be underestimated. Consequently, it is understandable why previous studies of citation behaviour have concentrated on detailed analysis within small areas of research or broad analysis of wider disciplines, rather than showing a detailed analysis of a comprehensive range of research as presented here.

### Dissemination behaviours

One important finding from this study was the number of authors who did not appear to have any identifiable publications in the years being analysed. By using RAE submission lists as the sampling frame, the population of authors from which the samples were drawn was expected to be 'research-active', and although some will inevitably be more active than others, around half of

those in the sample had not published outputs, which we could identify. This is a relatively high proportion, across all subject areas, and further investigations in this area would be of interest. Given this finding, it is perhaps surprising that the analysis found an increase in the average number of outputs per author overall, particularly in Bio-medicine and Social sciences. The reasons for this are not clear. One suggestion is that authors are beginning to prepare two or more shorter outputs where previously they may only have written a single, longer, one. There are insufficient data available in this analysis to test this hypothesis in detail by subject. However, there are some indications in our data that there were fewer journal articles and more abstracts and editorial material produced in Bio-medicine in 2008 than in 2003. There are increasing restrictions on word counts, both in terms of journal editors setting limits (Fry et al. 2009b) and because of increases in publication pressure within specific fields (Becher & Trowler 2001). A second factor contributing to the increase in outputs per author may be the rise in collaboration and multiple-authored papers (see for example de Villiers 1984, Madison 2003, Kyvik 2003, Drenth 1998). Increasing multiple-authorship may have resulted in individual authors being named on outputs to which they have made a relatively small contribution. Anecdotal evidence of this emerged from the focus group discussions undertaken as part of the wider research project (Fry et al. 2009b).

### **Citation practice**

Although there was no statistically significant difference in the average number of references per output between the two years examined, the observed mean for 2008 was, in fact, lower than in 2003. Others have found evidence of journal practices limiting the numbers of references allowed; for example, Harwood (2008) noted that restrictions on the word counts imposed for journal articles have had a direct effect on citation habits, both in using references to point to a detailed explanation of what is being summarised and by the careful and selective pruning of references to provide maximum impact and relevance in the space provided.

There were some interesting indications of interactions between subject and publication type when considering citation behaviour. It is only to be expected that monographs will include more references than journal articles, but there is some suggestion of disciplinary differences, particularly in the numbers of references included in monographs. There were insufficient data to draw firm conclusions in this area, however, and further investigations would be desirable to confirm the preliminary findings.

There was a greater proportion of outputs which did not cite any references in 2008 than in 2003. This is likely to be associated with the types of output, and there are indications that there were more editorial pieces, abstracts and book reviews in 2008 than in 2003. There were insufficient data to test this formally, however.

This study attempted to count websites as a separate category. It was difficult to distinguish between websites as an output in their own right and websites as a vehicle for other outputs. It was clear that the internet has become a widely used route to access research outputs, with many more instances of such use being seen in 2008 than in 2003, although, data were not collected to quantify this. The Social sciences show a greater use of websites in research, though the small proportion of these compared to other outputs cited shows that they remain relatively minor. Bio-medicine, Science and Engineering show almost no use of websites, although the abbreviation of such references by WoS makes them very hard to detect.

One aspect of this study that may be useful to other areas of bibliometric investigation is the collection of data that show the source material used by those subjects that are not considered to have good coverage in WoS. Moed (2005) shows overall ISI coverage of literature for fifteen broad subject disciplines, ranging from molecular biology & biochemistry with 92% coverage to humanities & arts with only 17% (the data are from 2002). Moed states that humanities & arts had an "importance of journals" of 34% and this corresponds to a similar number of references to journal material found in 2003 for Humanities.

### **Recommendations for future research**

A further study that looks at each year between 2003 and 2008, and beyond, would inform the interpretation of differences found, and allow more detailed examination of the effects of the research assessment cycle.

By using WoS, the method for obtaining lists of research outputs in the biomedical, physical science and engineering disciplines was biased towards journal articles, and uncovered only one book and no book chapters. The survey of researchers carried out as a parallel strand of the wider

project (Fry et al. 2009c) showed that monographs in particular are not considered particularly important in these disciplines, but they are used to maximise dissemination to particular target audiences. Although using WoS was an efficient data source, in any future similar studies, we suggest that it should be supplemented by other databases to discover a wider variety of outputs in these disciplines.

It would also be instructive to revise the method for any future studies of this nature regarding the classification of websites by including a non exclusive category for references with URLs in order to better reflect the use of the internet as an information source.

## **Acknowledgements**

The authors would like to thank the Research Information Network for funding the project reported here, the Expert Panel who oversaw the project, the reference classifiers, the administration staff at LISU, and the members of the wider project team: J. Fry, C.W.D. Johnson, and S. U. White (Loughborough University), and G. Butters, J. Craven, J. Griffiths, and R. J. Hartley (Manchester Metropolitan University).

Finally, we would like to thank the two anonymous reviewers for their helpful comments.

## References

- Becher, T. & Trowler, P.R. (2001). *Academic tribes and territories: intellectual enquiry and the cultures of disciplines*. 2nd revised ed. Buckingham: Open University Press.
- Broadus, R. N. (1971). The literature of the social sciences: a survey of citation studies. *International Social Sciences Journal*, 23, 236–243.
- Cronin, B., Snyder, H., Atkins, H. (1997), Comparative citation rankings of authors in monographic and journal literature: a study of sociology. *Journal of Documentation*, 53 : 263–273.
- Cullars, J. M. (1998), Citation characteristics of English-language monographs in philosophy. *Library & Information Science Research*, 20, 41–68.
- de Villiers, F.P.R. (1984). Publish or perish – the growing trend toward multiple authorship. *South African Medical Journal*, 66.
- Drenth, J.P.H. (1998). Multiple authorship: the contributions of senior authors. *The Journal of the American Medical Association*, 280, 219-221.
- Fry, J., Oppenheim, C., Creaser, C., Johnson, C.W.D., Summers, M.A.C., White, S.U., Butters, G., Craven, J., Griffiths, J. & Hartley, R.J. (2009). Publication, dissemination and citation behaviour of researchers: supporting paper 1: bibliometric analysis. <http://www.rin.ac.uk/system/files/attachments/sarah/Communicating-knowledge-bibliometric-analysis.pdf>. Accessed 19 February 2010.
- Fry, J., Oppenheim, C., Creaser, C., Johnson, C.W.D., Summers, M.A.C., White, S.U., Butters, G., Craven, J., Griffiths, J. & Hartley, R.J. (2009). Publication, dissemination and citation behaviour of researchers: supporting paper 2: report of focus group findings. <http://www.rin.ac.uk/system/files/attachments/sarah/Communicating-knowledge-focus-groups.pdf> Accessed 19 February 2010.
- Fry, J., Oppenheim, C., Creaser, C., Johnson, C.W.D., Summers, M.A.C., White, S.U., Butters, G., Craven, J., Griffiths, J. & Hartley, R.J. (2009). Publication, dissemination and citation behaviour of researchers: supporting paper 3: report and analysis of researcher survey. <http://www.rin.ac.uk/system/files/attachments/sarah/Communicating-knowledge-researcher-survey.pdf>. Accessed 19 February 2010.
- Glänzel, W. and Schoepflin, U. (1999). A bibliometric study of reference literature in the sciences and social sciences. *Information Processing and Management*, 35(1), 31–44.
- Harwood, N. (2008). Publication outlets and their effect on academic writers' citations. *Scientometrics*, 77(2), 253-265.
- HEFCE, SHEFC, HEFCW & DELNI ([n.d.]a). RAE 2008 [: panels]. Available at: <http://www.rae.ac.uk/panels/>. Accessed 19 February 2010.
- HEFCE, SHEFC, HEFCW & DELNI ([n.d.]b). RAE 2008: submissions. <http://www.rae.ac.uk/submissions/>. Accessed 19 February 2010.
- Kyvik, S. (2003). Changing trends in publishing behaviour among university faculty, 1980-2000. *Scientometrics*, 58(1), 35-48.
- Larivière, V., Archambault, E., Gingras, Y. & Vignola-Gagné, E. (2006). The place of serials in referencing practices: comparing natural sciences and engineering with social sciences and humanities. *Journal of the American Society for Information Science and Technology*, 57(8), 997–1004.

Linhholm-Romantschuk, Y., Warner, J. (1996), The role of monographs in scholarly communication: an empirical study of philosophy, sociology and economics. *Journal of Documentation*, 54, 389–404.

Madison, C.L. (2003). Review of published literature JSLHR: selected parameters. American Speech-Language-Hearing Association Convention, Chicago, November, 2003 (session 183).

Moed, H. (2005). *Citation analysis in research evaluation*. Dordrecht: Springer.

Nederhof, A. J. (1995), A bibliometric study of literature projects funded by the NWO Stichting Literatuurwetenschap. Leiden: report CWTS-95-05.

Nederhof, A.J. (2006). Bibliometric monitoring of research performance in the social sciences and the humanities: a review. *Scientometrics*, 66(1), 81-100.

Thompson, S.K. (2002). *Sampling*. 2nd ed. New York: Wiley.



Authors' response to reviewers' comments regarding "What do UK academics cite? An analysis of references cited in UK scholarly outputs".

Reviewer #1:

<b>Comment</b>	<b>Response</b>
1. Main findings are in literature	We do not understand this comment
2. Literature not reviewed well.	We were not attempting to produce a systematic review of the literature. However, we accept that more contextual material would be helpful so have re-written the introduction to include more relevant literature.
3. Unfavourable comparison of Moed and Meadows.	This was not meant to be read as such, rather as two 'snapshots' of information. This was not explained as well as it might be and this is reflected in the re-written introduction.
4. Sample of research active staff thought to be inaccurate.	Staff included in the RAE are considered by their institutions to be research active. Clarified.
5. Duplicate names dropped incorrectly.	Clarified that suspected duplicate people, not names, were removed prior to sampling.
6. Small increase in outputs over time possibly a consequence of sampled authors becoming more senior.	The samples were independent and there was only a very small likelihood of the same author appearing in both samples. Clarified.
7. "The increase in publication in social sciences and humanities is discussed already and the move in engineering from conference series to journals too is known."	This paper attempts to show the specific and comprehensive picture of UK. In this regard, our findings support those of other studies
8. Very long and not correctly structured.	The revised manuscript is shorter after further editing and corresponds structurally with previous <i>Scientometrics</i> articles.
9. Many tests could be in one table rather than in the discussion.	We do not see how this could be achieved
10. Possibly a big study for a thesis.	Yes, this was a part of a big study, but as stated in footnote 2, it was not for a thesis; it is important to have results published after peer review to share them with an interested audience.
11. Most readers of <i>Scientometrics</i> will not find value in such a long development of things they have already seen.	<i>Scientometrics</i> has previously published country-specific studies but we have not found another that deals with issue either in terms of the UK or involved the use of a novel method on such a comprehensively broad scale in this way.

Reviewer #2:

<b><i>Comment</i></b>	<b><i>Response</i></b>
i) Suggested removing sections on collaboration as they are distracting	Sections removed from results and discussion.
ii) Suggested removing redundant detail.	Further editing undertaken throughout the paper.
iii) Queried the inclusion of some publication types as research outputs.	We felt that meeting abstracts were valid research outputs as in certain disciplines these represent important dissemination channels for the results of research (e.g., in computer science). Also, our intention was to record the entire breadth of researchers' outputs and to start excluding certain types would involve both subjective decisions and start to eat away at the paper's comprehensive overview.
iv) Charts and tables unclear.	Redrawn – if the editors have any specific suggestions for greater clarity, please let us know.
a), b), c) specific comments on the introduction.	Introduction entirely rewritten – see above
d) HESA should be spelt out.	Changed
e) Suggested removing last column of Table 1.	Removed
f) suggests adding indication of the problem in the Arts.	Added a reference to the lack of data. More discussion of this point can be found later in the article
g) did we use the RAE database?	Yes, this was looked at, though very few relevant papers were found using this source. List amended.
h) why are items “in press/ forthcoming” a separate category?	We felt that items in press/forthcoming have not been verified by publication so cannot be relied upon as existing as publications, therefore should not be added to the category to which they appear to belong. Clarified.
i) need for more information on weighting.	Added
j) Chart difficult to read.	Redrawn
k) Font size in figure 2.	Altered
l) Inconsistent headings on table 8.	Made more consistent
m) collaboration section.	Removed
n) suggested further table on dissemination behaviours.	Not clear how we could do this
o) meeting abstracts in the RAE.	On reflection, the point made here about RAE panels seems to be unimportant without a much more in-depth discussion, and is therefore probably inappropriate here, so has been removed.