



This article has been submitted to Loughborough University's Institutional
Repository by the author.

**Citation counts and the Research Assessment Exercise V [1]:
Archaeology and the 2001 RAE**

Michael Norris and Charles Oppenheim*
Department of Information Science
Loughborough University
Loughborough
Leics LE11 3TU
UK

C.Oppenheim@lboro.ac.uk

Word count 7,564 words

Abstract

A citation study of the 692 staff that makes up Unit of Assessment 58 (Archaeology), in the 2001 UK Research Assessment Exercise (RAE) was undertaken. Unlike earlier studies, which were obliged to make assumptions on who and what had been submitted for assessment, these were, for the first time available from the RAE Web site. This study, therefore, used the specific submission details of authors and their publications. Using the Spearman Rank Order Correlation Coefficient, all results showed high statistically significant correlation between the RAE result and citation counts. The results were significant at 0.01%. The findings confirm earlier studies. Given the comparative cost and ease of citation analysis, it is recommended that, correctly applied, it should be the initial tool of assessment for the RAE. Panel members would then exercise their judgement and skill to confirm final rankings.

Keywords: CITATION COUNT; CITATION ANALYSIS; ARCHAEOLOGY; RAE; RESEARCH ASSESSMENT EXERCISE

*To whom correspondence should be addressed

Introduction

In this paper, we describe research on the correlation between citation counts and an official 2001 assessment of research excellence in UK University archaeology departments. To assess the research performance of UK universities, the UK Higher Education Funding Councils carry out a periodic Research Assessment Exercise (RAE). The RAE is concerned with making a qualitative judgement of the research output of those university departments who submit themselves to the procedure. As part of this assessment, academic (and in some cases industrial) peers carry out a review of the published work of submitting departments over a given period of time. The departments are also assessed by other criteria, including the number of higher degree students they have, the amount of research income they have received and a general commentary on their current and future research programmes. University departments are then ranked and consequently funded by central government for their research activities, based upon the score they achieve.

The RAE is carried out every five years or so and is viewed by many as an expensive and contentious process. Alternatives are sometimes suggested that are claimed to be both cheaper and less divisive. One such alternative is citation analysis.

Generally, the more citations an author receives, the more important that author's work is assumed to be. (Baird & Oppenheim, 1994) Previous research (see below) has shown that there is a strong correlation between the ranked scores obtained by the RAE and the ranked scores found by counting the citations received by authors in their research writings. Given the considerable amounts of money involved in funding research in higher education, the process of allocating funds to any research institution is likely to be contentious. The RAE has been the mechanism, since 1986, by which the research funding apportionment has been resolved. The RAE process itself has generated a large amount of interest. It has itself become, ironically, a subject of research and a significant amount of

material has been published on its effects and the possible alternatives that may be used in its place. Henkel provides a good overview of the background and effect of the RAE on universities (Henkel, 1999). Martin gives an assessment of multiple performance indicators that are available to measure research output (Martin, 1996). Amongst the alternative performance indicators that can be used to measure research output is citation analysis. This could be used to assess an academic department in terms of counting the number of citations that its staff members have received for their published work. Such a set of statistics from each department could then be used as an alternative funding hierarchy (Holmes & Oppenheim, 2001).

The RAE affects not only UK Higher Education Institutions, but also the people involved in its management and those academics whose research output is critically judged. The effect of the RAE extends well beyond the boundaries of academia and government. The scholarly publishing sector is also noticeably affected by it. Similarly, those deciding where to place research funding and those simply looking to find the best place to do research will be influenced by RAE results. At a time of growing international competition, the cost of basic research rising and the continuing need to demonstrate the usefulness of university research, the first Research Assessment Exercise was instituted in 1986. Including the 2001 RAE, there have been five such assessments.

The 2001 Research Assessment Exercise

The 'Guide to the 2001 Research Assessment Exercise' (Publications, 2002) gives an appreciation of the scale, history and importance of the 2001 RAE. The 'Guide' states:

The purpose of the RAE is not just to enable funding to be allocated selectively but also to promote high quality: research in higher education institutions conducting the best research receive the largest proportion of the grant.

Funding decisions are about promoting high quality research and improving the productivity and effectiveness of that research. A document from HEFCE reports the findings of a consultation process on research (HEFCE, 2001b). It claimed an overall satisfaction level of 98% with the question:

Should there continue to be a research assessment process based on peer review, building on the foundations of the RAE?

The respondents were made up of 102 English universities and colleges: 32 subject associations and 30 other organisations. All were in broad agreement with the overall thrust of the funding strategy. The closing date for consultation was, however, December 2000, a year prior to the announcement of the final RAE results.

Rogers (at the time responsible for the RAE) disagrees with the many criticisms of the RAE. He argues, for example, that there is no wholesale movement of staff, nor is interdisciplinary research discriminated against and that the diversity of published forms that are acceptable as submissions is wide and all are viewed as being equally valid (Rogers, 2000).

For the individual, the effects of the RAE may be extreme. Individuals designated as 'research inactive' may find that they have a drastic loss of status and are relegated to higher administrative and teaching loads. The individuals may be viewed as liabilities. 'Research active' staff are likely to find themselves being managed and directed in an atmosphere of 'collective interest rather than individual ambition, and structures and policies have been developed accordingly' (Henkel, 1999).

Scholarly Publication and the Research Assessment Exercise

Several surveys and much anecdotal evidence suggest that the RAE and its demands on academics have affected the scholarly publishing industry. The RAE officially does not favour any one particular form of

material submitted to it for evaluation, yet clearly the high impact factor journal article will have the stamp of peer approval and a general recognition of quality already attached to it (Naylor, 2001). Campbell et al examined the affect that the RAE has had on publishing with particular regard to academic lawyers (Campbell, 1999). The clear view emerges that journal articles are important and the actual journal in which articles appear will lend authority to the work published. It is widely (but probably incorrectly) believed that the RAE assessment panel will view such work more favourably than more obscure publications. This conclusion is not shared by Bence and Oppenheim who examined the field of business and management studies (Bence & Oppenheim, 2001). The subject area attracted in the 1996 RAE 5494 articles to 1275 different journal titles. They concluded that the academic author in this field attaches importance to the journal in which their work appears, but that there is a much more diverse range of journals and this makes the task of defining a core set of high impact factor journals much more difficult.

In the humanities, there is a noticeable preference for publishing academic research in monograph form, unlike scientific research. Meadows suggests this difference may be seen as the 'hard' quantitative nature of scientific subjects being suited to journal publication and the more discursive 'soft' humanities research lending itself to monographic publication (Meadows, 1998). In the publishing industry itself, Mynott of the Cambridge University Press clearly identifies the pressures placed upon him. Some of the normal editing and revision processes were foreshortened and there was a 'competition' with some works which would only be offered if they could be published within the given time-scale (Mynott, 1999). He noted that his Press published 122 more titles in the first six months of 1996 compared to the previous year.

Walford also found a sharp increase in the number of contributions to scholarly journals in the run up to a RAE (Walford, 2000). He suggested that whilst some

‘salami-slicing’ took place, there was also a proliferation of run-of-the-mill papers, which, though weeded out by the peer review process, put strain on the refereeing process.

A new discovery in medicine confers status to the published work and the author, therefore making it excellent RAE material. The humanities scholar’s work, on the other hand, has to be judged much more finely on the contribution of its scholarship to the body of existing knowledge and this, Bernard argues, makes judging the quality of the work much more difficult (Bernard, 2000)..

Complexities start to arise in the humanities when scholars begin to interpret their research material in controversial ways. Whilst scientific discoveries can be tested and experiments replicated, the opinion of one historian against another cannot be tested to verify its truth. Thus, the quality of some particular research as well as its suitability for RAE submission could be questioned. It could be argued that archaeology is in a similar position to historical research, since although it is based on clear evidence (archaeological dig findings; historical records), much of the work is speculative and based on the judgement and interpretation of the researcher. Such possible ambiguities in assessing the quality of humanities research clearly allow for the possibility of inaccuracies in awarding RAE ratings.

Citation Analysis

Citation analysis is based on counts of the number of citations that an article or monograph has received in a particular period. Citation analysis was not easily undertaken until the launch of *Science Citation Index*, *Social Sciences Citation Index* and *Art and Humanities Citation Index*. Garfield’s seminal work on citation analysis, although 25 years old, provides an authoritative historical and conceptual understanding of the topic and its many applications (Garfield, 1979). A more up to date overview can be obtained from (Cronin & Barsky Atkins, 2000).

The use of citation studies in place of, or as a supplement to, the RAE assessment has been considered extensively by Oppenheim and his co-workers. They have published a number of articles demonstrating that the correlation between the overall RAE ranking of departments and the collective scores found by counting the citations received by the authors from those departments is statistically significant (Holmes & Oppenheim, 2001; Oppenheim, 1995; Oppenheim, 1996, 1997). He argues that citation analysis costs a fraction of the RAE and is much less intrusive than the RAE. This has been followed up with related suggestions for improving the RAE in the future (Harnad, Carr, Brody, & Oppenheim, 2003).

Other researchers have found similar strong correlations between citation counts and RAE scores. Seng and Willett examined the correlation between the publications of UK library schools between the years 1989-1990 and the 1992 RAE result (Seng & Willett, 1995). More recent work to consider citation analysis as an alternative or supplement to the RAE was carried out by Smith and Eysenck, who examined the citation record of psychologists for the year 1998; citations were checked for their probity to eliminate self-citations and correct any misnamed individuals. The results were compared to the 1996 and 2001 RAEs. A 0.9 and 0.85 correlation was found for all-author citations respectively (Smith & Eysenck, 2002). These statistically significant results, they argue, are ample supporting evidence for the use of citation analysis as an effective evaluation tool of research quality. Sarwar studied UK Civil Engineering departments to establish if there was a correlation between the 1996 RAE and a citation analysis of those departments who had achieved a rating of 4 or above. Comparing the citation record of the authors in those departments for the years 1995-97 and the department's 1996 RAE rating, he (Sarwar, 2000) confirmed a significant correlation between the two sets of results.

In one of his papers, Oppenheim examined the subject area of archaeology (Oppenheim, 1997). He demonstrated that archaeology produced the highest correlation of the three subject areas studied.

Citation analysis has many critics, and some have focussed their criticism on the idea that citation analysis could be used as an input into the RAE process (Johnson, 2001; Warner, 2000), or on the bias introduced by self-citation (Seglen, 1992). The response of proponents of the method is the inherent robustness of citation analysis as a valid tool of measurement since all the experiments so far have shown strong statistically significant correlations with RAE results. The question of self-citation is dealt with satisfactorily by (Snyder & Bonzi, 1998).

Whilst Oppenheim is a strong proponent of the use of citation analysis to establish a funding hierarchy, he has concluded (Baird & Oppenheim, 1994) that

[T]here is not, and never can be one single measure of the value of information that will be universally acceptable. However, there are a number of measures that might, in combination, lead to some sort of index of the value of a piece of information, an individual's research contribution, or a collection of information.

This collection view of 'measures', which together could be used to provide a more balanced and objective indicator of research quality is one shared by Martin (Martin, 1996) who analysed the possible measures that could be used. Martin crucially notes that citation counts are an indicator of impact of the author, their past reputation and the organisation in which they work. He concludes that high citation counts may not always indicate quality or importance, but may be for controversy, fraud or a simply incorrect work. Studies underway in this Department [2] are examining citation counts for controversial articles. Martin also makes the important observation that the more indicators of research you have, the more difficult it is for individuals to manipulate them to their advantage.

Van Raan describes (Van Raan, 1996) an assessment exercise undertaken in the Netherlands, where peer review is used jointly with advanced bibliometric techniques to evaluate research. The majority of the reviewers come from

outside the Netherlands and they undertake a quality assessment of the research output from the universities. In addition, a scrupulous bibliometric assessment process designed to remove the many anomalies found within citation counting such as self-citing, incorrect counts, journal coverage, etc., was carried out. The result is a 'cleaned' citation count. . Van Raan concluded that:

We showed that the resulting indicators are useful: they address significant concepts in the framework of evaluation and therefore can be considered an indispensable element next to peer review in research performance assessment procedures.

This approach validates the use of citation analysis as a sound measure of research performance when used along side peer review.

The research described in this paper was aimed at establishing whether the well-established correlations between RAE performance and citation counts continue to apply to the 2001 RAE, and also whether the ability to more precisely identify who had been returned for the 2001 RAE leads to a better correlation than simply carrying out an analysis based on all staff in a Department. Archaeology was chosen as a subject for study for two reasons: firstly, it had been studied before, so comparisons with earlier results could be made; and secondly, it is not such a large discipline, so the research could be completed in a reasonable time frame.

Methods Employed

A citation analysis was undertaken which counted the citations received by those academics in archaeology departments submitted for peer review in the 2001 RAE. In earlier assessments, it was not possible to identify the specific academics returned by any particular department, nor the publications they submitted for assessment. This is now, however, possible for the 2001 RAE. Thus the methods adopted differ slightly from previous similar studies. It is worth stressing that each academic's submission is recorded on the RAE Web site, and is therefore a public document. Therefore, no permission was needed

from this individuals to include them in our analysis, or to present results relating to them (see Table 8 below).

Archaeology was chosen following an assessment carried out to find a subject that was both sufficiently discrete to allow its careful measurement and was of the right size and scale that it could be comfortably managed. Archaeology is denoted in the 2001 RAE as Unit of Assessment (UoA) 58. Twenty-six university departments returned submissions into UoA 58. They were ranked in the 2001 RAE with scores from 3b to 5* for their research excellence. As noted above, the report of the 2001 RAE (HEFCE, 2001a) for the first time made available details of those academics that had been submitted for peer review by departments. All the names of the academics who were submitted for UoA 58 were thus obtained. Each academic was allowed to submit up to four items, which they had produced during the assessment period. . We noted details of all 682 staff, including name, institution and the details of the items submitted for the RAE. *ISI Web of Science* was used to carry out the citation counting. The searches were confined to the *Art and Humanities Citation Index* (AHCI), since it covers the field of archaeology. The cited reference searches were carried out between August and September 2002. The searches were confined to the RAE assessment period, i.e., 1994 to 2000 and looked for citations to material that was published during this interval by the authors concerned, noting the number of citations this material had received. Named authors were normally entered into the search as given. In the case of hyphenated or compound names, standard ISI practice of contraction was used. For example the name Van Der Leeum was entered as Vanderleeum. Some difficulty was encountered with authors who had several initials. Where the search returned a significant number of hits for all the initials, it was considered that the author routinely used their fully initialled name in authorship. Where a nil or very low return was encountered for a multiple initialled author, further searches were carried out by progressively reducing the number of initials until an appropriate result was yielded. The cited references were first scrutinised to ensure that they fell within the RAE assessment period, i.e., 1994 to 2000. Each cited reference was then examined to ensure it was the right subject and that it could be attributed to the author concerned. Use was

made of the author's submission to the RAE, which indicated his or her subject area. Where ambiguity was apparent in authorship, records were marked and the abstract and cited references examined to verify that the correct author and subject had been identified. Similarly, ambiguous records, which had direct electronic links to an abstract and cited references, were also scrutinised.

A small number of cited works were ignored, i.e., those listed without publication dates, unpublished works and those 'in press'. Where ambiguity was still apparent in attributing authorship, a search was made at the author's departmental web site and access to their publication record clarified responsibility for the cited work.

Each author's citation count was then calculated. Anomalous records were rechecked as necessary. The 682 named authors were transcribed onto a spreadsheet into their respective departments and citation counts attributable to them were likewise entered. The authors were differentiated to show which of them had been submitted for peer review. Of the 682 authors listed, 73 had not been submitted for review. From this general data, a number of specific citation statistics by department, including averages, were calculated to allow subsequent correlation calculations to be made.

In order to carry out a Spearman Rank Order correlation calculation (the method used in all these studies), we ranked the archaeology departments by their RAE scores and their different citation statistics. The twenty-six archaeology departments were listed in ascending order of their 2001 RAE ratings. Each department was then assigned a separate ascending numerical rank. Where there were two or more equal values in the departmental rating, an average of the numerical ranking was applied. For example, Reading, Oxford and Cambridge each had a 5* (the highest possible) rating and so each were given a 2 ranking.

The citations from the authors were collated into their respective university departments and totalled. From these collations, a set of four different tables were produced each containing the twenty-six departments, their RAE rating and

the total number of citations. Two of the tables were differentiated by a calculated average citation rate dependent on the status of the cited author. The other two tables were differentiated by the total citations received by each department but again dependent on the status of the cited authors. In both sets of tables, the criteria for differentiation were submitted and non-submitted authors and their citation scores. To again obtain a Spearman Rank Order correlation each department in each table was then assigned a numerical rank and, if necessary, an average numerical rank applied.

Four sets of calculations were then performed:

- A ranking for all citations received by each department, irrespective of whether the author had been submitted or not
- A ranking by the total citations received by each department counting submitted authors only
- A ranking which shows the average citation rate for each department using all departmental staff, irrespective of whether the author had been submitted or not and
- A ranking, which shows the average citation rate by department using, submitted departmental staff only.

Issues Arising Out of the method adopted

Citation counts, can, however well carried out, lead to erroneous counts. The methods adopted in this study go a long way to minimise such erroneous counts. A number of commonly voiced criticisms of citation counting (Liu, 1993; MacRoberts & MacRoberts, 1989) were potentially present in this study and are worthy of comment. The general problem of mistakenly counting authors with the same surname and initials, but in different disciplines, was understood and

careful inspection was exercised to limit such miscounting. This process was considerably helped by the RAE submissions, which identified the authors' interests and some of their publications. Examination of the bibliographical record of the citing article for its subject and its fellow citations helped minimise potential error. Unquestionably, however, judgement was required. Where it was very difficult to assess the validity of a particular citation, the record was included. Such cases were infrequent and are not likely to distort the total counts or their relative rankings. In any case, it is the comparative ranking of the results, which is being sought, rather than an absolutely accurate citation count.

Where formerly only first named authors were cited, now second and subsequent authors are also cited in the ISI's indexes, thus, the comment that co-authors were uncited within the index and hence uncounted no longer applies. This has improved the citation count achieved and removed a significant objection. However, Lange suggests that some caution needs to be exercised in assuming that a first and subsequent cited author search can be deemed to be absolutely exact (Lange, 2001). Self-citation has long been criticised in citation analysis; such practice, it has been suggested, leads to distortion and bias in conclusions drawn from counts. The evidence and studies conducted so far, however, show that this practice has little effect on the results of citation analysis. Snyder and Bonzi, in a study which examined the patterns of self-citation in six disciplines, clearly showed its prevalence, but were able to demonstrate that the patterns of self-citation are very similar between disciplines and that the lowest self-citation rate of 3% was found in the humanities, including presumably archaeology (Snyder & Bonzi, 1998).

An author making a submission to the RAE may choose material that has been published at any time during the qualifying period. This may make the items submitted potentially 'old' or very 'new' material. Counting the citations to a population of 'new' articles against 'old' could easily give a skewed result in favour of the old articles, given that more citations are likely to accrue to the older material. We did not examine this in detail, but a simple overview indicated

that the pattern of age of items submitted did not differ much between the departments examined.

Humanities scholars tend to publish more monographs than those in other disciplines. A random sample of 40 submitted authors examined here, credited at least 32 of them with either having published a monograph or chapter in one submitted for peer review. In many cases, they had published more than one monograph or chapter in the assessment period. Cronin *et al* have examined the differing author citation patterns when comparing journals and monographs in the field of sociology (Cronin, Snyder, & Atkins, 1997). They suggest that journals and monographs attract two quite distinct populations of authors amongst the very highly cited: this is certainly a concern, and it can be assumed that insufficient citations totals will result from this. However, we had no evidence that this affected one department more than another, and it is worth reiterating that it is the comparative ranking of citation counts that is studied here, rather than the precise count.

A criticism may be levelled at the scope and coverage of the AHCI, in terms of its bias towards USA/European publications as a distorting feature when making citation counts (Braun, Glanzel, & Schubert, 2000). Therefore, where an author has made contributions to non-USA/European literature, there is less likelihood of his or her work being noted. V. Nanda of University College London did not receive any citations for his work, which has been published in Asia. Likewise, P.M. Carroll of the University of Sheffield did not receive any citations to his works when published in German. Whilst this is not conclusive evidence as to any particular bias, the work of Braun *et al* clearly shows a strong preference for a USA/European dominance of journal titles and publishers within *Web of Science*. Examining the Archaeology Journal List in the AHCI confirms a preference for USA/European journals. Table 1 below shows the distribution of journal title by place of publication.

**Table 1. Distribution of Journals Indexed by ISI Web of Science
for Archaeology**

Number of Journals	Place of Publication
Denmark	1
France	4
Germany	3
Israel	1
South Africa	1
UK	10
USA	17
Total	37

This shows there is a definite bias toward USA/European journals, so the likelihood of finding the cited work of UK archaeologists is considerable – unless they publish elsewhere. This bias, in practice, adds credibility to this particular study. This, it is contended, will help support the results obtained and the conclusions drawn.

Results

The results have been summarised in a number of Tables. These Tables give several views of the data, notably by a ranking of the RAE score obtained by each archaeology department and several different citation statistics. **All** correlations found were significant, thus confirming earlier studies. We also found that correlations for just staff returned for assessment is even higher, thereby arguing that this is an even better predictor of RAE results.

Table 2 gives an overall view of the RAE score obtained by each department and a comparative ranking by the four basic citation statistics that have been calculated for each of them. Table 3 describes the core statistics collected and of the range of values obtained for each of the departments. Calculations using the core data and the RAE score give the four measures of correlation required. Two of the measures deal with total citation counts and the other two are concerned

with average citations per member of staff. The 2001 RAE rating achieved by each archaeology department and their respective citation statistics are shown in Table 4. The Table has been sorted within each RAE rating by the number of submitted staff citations received, giving a snapshot view of the range of citations received and also how this compares between ratings. To compare the 1996 and 2001 RAE ratings, Table 5 shows the scores obtained in both assessments and the movement between the two periods.

Table 2. Comparative ranking with the RAE Rating Result 2001, Staff Citation Counts, and Citation Count Averages for UoA 58 Archaeology

University Name	RAE result 2001	Ranking by submitted staff citations	Ranking by all citations	Ranking by avg citations for submitted staff only	Ranking by avg citations for all staff members
Trinity College Carmarthen	3b	25	23	20	24
King Alfred's College Winchester	3a	24	24	26	25
University of Wales College Newport	3a	23	25	25	26
University of Edinburgh	3a	21	21	22	21
Bournemouth University	3a	18	18	21	20
University of Newcastle	3a	17	17	17	17
University of York	3a	15	15	18	18
Birkbeck College (one staff member only)	4	26	26	12	13
University of Bristol	4	22	22	24	23
University of Nottingham	4	19	19	15	15
University of Birmingham	4	16	16	23	22
University of Glasgow	4	14	14	19	19
University of Wales Lampeter	4	13	11	13	9
University of Exeter	5	20	20	16	16
University of Liverpool	5	12	13	11	11
Cardiff University	5	11	12	7	8
University of Leicester	5	9	9	14	14
University of Southampton	5	8	8	8	7
University of Bradford	5	7	6	5	4
University of Durham	5	6	7	9	10
University of Sheffield	5	5	5	4	3
University College London	5	4	4	10	12
The Queens University of Belfast	5	2	2	1	1
University of Reading	5*	10	10	6	6
University of Cambridge	5*	3	3	3	5
University of Oxford	5*	1	1	1	2

Table 3. Summary Results of Citation Counts, Staff Numbers and Citation Count Averages for UoA 58 Archaeology

University Name	Submitted staff citations	Non-submitted staff citations	Total all Citations	Total submitted staff	Total non-submitted staff	Avg citations per submitted staff member	Avg citations per non submitted staff member	Avg citations all staff members
Birkbeck College	6	0	6	1	0	6.00	0.00	6.00
University of Birmingham	66	0	66	30	2	2.20	0.00	2.06
Bournemouth University	47	0	47	16	0	2.94	0.00	2.94
University of Bradford	238	78	316	19	3	12.53	26.00	14.36
University of Bristol	37	0	37	18	0	2.06	0.00	2.06
University of Cambridge	811	88	899	43	23	18.86	3.83	13.62
University of Durham	252	2	254	36	3	7.00	0.67	6.51
University of Exeter	41	0	41	11	0	3.73	0.00	3.73
King Alfred's College Winchester	11	5	16	7	2	1.57	2.50	1.78
University of Leicester	166	3	169	35	2	4.74	1.50	4.57
University of Liverpool	111	0	111	18	0	6.17	0.00	6.17
University of Newcastle	51	6	57	15	2	3.40	3.00	3.35
University of Nottingham	45	0	45	12	0	3.75	0.00	3.75
University of Oxford	1192	76	1268	59	11	20.20	6.91	18.11
University of Reading	164	0	164	20	0	8.20	0.00	8.20
University of Sheffield	314	62	376	20	2	15.70	31.00	17.09
University of Southampton	224	23	247	30	3	7.47	7.67	7.48
University College London	550	9	559	88	4	6.25	2.25	6.08
University of York	77	0	77	23	0	3.35	0.00	3.35
University of Edinburgh	38	0	38	14	0	2.71	0.00	2.71
University of Glasgow	103	0	103	31	0	3.32	0.00	3.32
Trinity College Carmarthen	6	12	18	2	7	3.00	1.71	2.00
Cardiff University	143	0	143	19	1	7.53	0.00	7.15
University of Wales Lampeter	106	53	159	19	5	5.58	10.60	6.63
University of Wales College Newport	14	0	14	8	0	1.75	0.00	1.75
The Queens University of Belfast	977	6	983	15	3	65.13	2.00	54.61
Grand Totals	5790	423	6213	609	73	9.51	5.79	9.11

Table 4. Results with Ranked RAE Rating, Citation Counts and Citation Count Averages for UoA 58 Archaeology

University Name	RAE result 2001	Submitted staff citations	Non- submitted staff citations	Total all Citations	Avg citations per submitted staff member	Avg citations per non submitted staff member	Avg citations all staff members
Trinity College Carmarthen	3b	6	12	18	3.0	1.7	2.0
King Alfred's College Winchester	3a	11	5	16	1.6	2.5	1.8
University of Wales College Newport	3a	14	0	14	1.8	0.0	1.8
University of Edinburgh	3a	38	0	38	2.7	0.0	2.7
Bournemouth University	3a	47	0	47	2.9	0.0	2.9
University of Newcastle	3a	51	6	57	3.4	3.0	3.4
University of York	3a	77	0	77	3.3	0.0	3.3
Birkbeck College	4	6	0	6	6.0	0.0	6.0
University of Bristol	4	37	0	37	2.1	0.0	2.1
University of Nottingham	4	45	0	45	3.8	0.0	3.8
University of Birmingham	4	66	0	66	2.2	0.0	2.1
University of Glasgow	4	103	0	103	3.3	0.0	3.3
University of Wales Lampeter	4	106	53	159	5.6	10.6	6.6
University of Exeter	5	41	0	41	3.7	0.0	3.7
University of Liverpool	5	111	0	111	6.2	0.0	6.2
Cardiff University	5	143	0	143	7.5	0.0	7.2
University of Leicester	5	166	3	169	4.7	1.5	4.6
University of Southampton	5	224	23	247	7.5	7.7	7.5
University of Bradford	5	238	78	316	12.5	26.0	14.4
University of Durham	5	252	2	254	7.0	0.7	6.5
University of Sheffield	5	314	62	376	15.7	31.0	17.1
University College London	5	550	9	559	6.3	2.3	6.1
The Queens University of Belfast	5	977	6	983	65.1	2.0	54.6
University of Reading	5*	164	0	164	8.2	0.0	8.2
University of Cambridge	5*	811	88	899	18.9	3.8	13.6
University of Oxford	5*	1192	76	1268	20.2	6.9	18.1

Table 5. A Comparison of the Ratings Achieved Over the Last Two Assessments

University Name	RAE Result 2001	RAE Result 1996	Rating Movement
Trinty College Carmarthen	3b	3b	0
King Alfred's College Winchester	3a	2	+
University of Wales College Newport	3a	3b	+
University of Edinburgh	3a	4	-
Bournemouth University	3a	3b	+
University of Newcastle	3a	3b	+
University of York	3a	4	-
Birkbeck College (one staff member only)	4	n/a	n/a
University of Bristol	4	4	0
University of Nottingham	4	3a	+
University of Birmingham	4	4	0
University of Glasgow	4	4	0
University of Wales Lampeter	4	3a	+
University of Exeter	5	3a	+
University of Liverpool	5	4	+
Cardiff University	5	4	+
University of Leicester	5	5	0
University of Southampton	5	5	0
University of Bradford	5	5	0
University of Durham	5	5	0
University of Sheffield	5	5*	-
University College London	5	5	0
The Queens University of Belfast	5	5	0
University of Reading	5*	5	+
University of Cambridge	5*	5*	0
University of Oxford	5*	5*	0

Table 6 summarises the results obtained for the Spearman Rank Order Correlation analyses for the four differing citation measures that were used. The correlation scores have been rounded to two significant figures.

**Table 6. Spearman Rank Order Correlation Coefficient for the 2001 RAE
Ranking and Citation Counts**

Statistic	Correlation
Total Submitted Staff Citations	0.81
Total of all Staff Citations	0.79
Total Staff Avg Citations	0.85
Total of all Staff Citations	0.84

Each of these correlation scores is highly significant ($p < 0.01$).

The results obtained by Oppenheim in his earlier study on archaeology are shown in Table 7 below for comparison. They are consistent with the current results.

**Table 7. Spearman Rank Order Correlation Coefficient for the 1992 RAE
Ranking and Citation Counts for Archaeology (Oppenheim)**

Statistic	Correlation
Total of all Staff Citations	0.82
Total Staff Avg Citations	0.74

Analysis of Results

682 staff members were checked for their citation counts. Between them, they accumulated 6213 citations, although 262 of the authors did not receive any citations at all in the assessment period. The individual citation count ranged from 0 to 565, with an overall average of 9.11. Of the cited authors, just over 50% of them have a citation level of between one and six citations each. The exceptional score of 565 citations obtained by P J Reimer can be attributed to just three articles. These were articles dealing with radiocarbon dating linked to Dendrochronolgy, and were as follows:

- *Quaternary Science Reviews*, 1996, 15(7), 655. **10** citations
- *Radiocarbon*, 1998, 40(3), 1041. **491** citations
- *Radiocarbon*, 1998, 40(3), 1127. **64** citations

Reimer's nearest rival was P Forster from Cambridge University, with 471 citations from four articles. They deal with DNA studies tracing the migration of humans from Africa:

- *American Journal of Human Genetics*, 1996, 59(4), 935. **108** citations
- *American Journal of Human Genetics*, 1996, 61(3), 691. **73** citations
- *American Journal of Human Genetics*, 1996, 59(1), 185. **164** citations
- *Genetics*, 1995, 141(2), 743. **126** citations.

These two authors account for one sixth of all the citations counted and their articles appear in just four journals, all of which are available in both print and electronic form.

Taking the top ten most cited authors and their collective citation counts shows them to account for 1993 of the totals citations received, i.e., 32% of the 6213 citations received by all authors. Table 8 below gives the details of the top ten most cited authors and their affiliations.

Table 8. The Top Ten Most Cited Authors and Their Affiliation

Author	Citation Count	Affiliation
P J Reimer	565	Belfast
P Forster	471	Cambridge
R E M Hedges	268	Oxford
E J Rhodes	127	Oxford
M G L Baillie	120	Belfast
M B Roberts	94	UCL
J Elsner	94	Oxford
J Pollard	92	Bradford
F G McCormac	81	Belfast
I R Hodder	81	Cambridge

Birkbeck College obtained the lowest citation score of six, for just one staff member. This low score would, have by citation count alone, placed it at the bottom of the ranking. He did, however, achieve a four rating by assessment and by average citation count here would have been rated as a borderline four or five. Both Bradford and Lampeter Universities did not submit all of their staff for assessment. The citation counts for those not submitted was a significant percentage (6.8%) of the total citation count achieved. The Table below shows those universities with authors with a significant number of citation counts who were not submitted.

Table 9. Percentage of non-submitted citations.

University	All Citations	Not submitted citations	Percentage
Carmarthen	18	12	67%
Lampeter	159	53	33%
Bradford	316	78	25%
Sheffield	376	62	16%

Carmarthen retained its rating of 3b with a citation count of 6. In the case of Lampeter, their ranking by total citation count would have placed them with a probable RAE rating of 5 against their current 4, however they did move from a 3a to the 4 rating in the current exercise. Bradford would have remained unchanged at 5. Perhaps the most significant change however, occurred for Sheffield who lost their previous 5* rating. Had Sheffield submitted K J Edwards, who accounted for 58 of the non-submitted citations and incidentally the highest citation count of all of their staff, then perhaps they may have retained their 5* rating. Given the relative simplicity and accuracy of making these ranking assessments by citation counting, the technique could be readily used as a guide to help decision making in borderline cases. Oppenheim has suggested (Oppenheim, 1996) that RAE co-ordinators would do well to consider a citation analysis of their authors before selecting them for submission. Proposals for the next round of assessment in 2007/2008 suggest a staged approach which would give Universities much more control of the process, and which would allow panels to select whatever methods of assessment they felt were most appropriate for their subject area (Roberts, 2003).

The RAE submissions included 2342 separate items made up as shown in Table 10.

Table 10. Breakdown of the Items Submitted for 2001 RAE Assessment

Items Submitted for Assessment RAE 2001 and Percentage Split					
Monograph	Conference	Reports	Internet	Articles	Other
1273	121	4	12	927	5
54.36%	5.17%	0.17%	0.51%	39.58%	0.21%

The 927 articles were submitted to 341 different journal titles with numbers ranging from a single submission to a number of journals through to 72 articles published in *Antiquity*, a UK based journal. Of the 37 archaeology journals indexed by *Web of Science*, 23 had articles submitted to them from amongst those presented for assessment. In total, the number of articles submitted to the 23 ISI journals were 203, 22% of the total number of articles presented. The 23 journals represent only 6.7% of the journals represented in the submissions.

When 5 and 5* rated Departments are considered separately, it is evident that the concentration of publishing authors of ISI indexed journals is greater than the lower rated universities. Table 11 shows that there is a relationship between the RAE rating, and the number of ISI indexed archaeology journals that authors contribute to from each of the Departments. Given that the journals selected by the ISI are high impact journals, it is not surprising that they attract the attention of leading authors.

Ten universities improved their score, twelve maintained their position and three lost a grade. There was also one new entrant making up the total of twenty-six departments. This is in line with the general ‘rating inflation’ experienced in the 2001 RAE.

Overall, the rankings obtained by citation analysis correlate very strongly with the rankings achieved by the RAE assessment. The results obtained are virtually identical to those found by Oppenheim on the previous RAE. This consistency of results over the two RAEs using an almost identical method supports the notion that

citation counting is a robust and reliable method, for making an initial ranking assessment of Archaeology Departments.

Discussion

Earlier citation studies that attempted to find a correlation between citation counts and the awarded RAE rating had to do so without knowing which academics had been submitted for assessment. The researchers in these studies gathered a listing of academics that they thought would be included in the assessment, and then carried out a citation analysis of them. The results from these studies showed a high correlation between the citation count of the academics in a particular department and the RAE rating. The citation study undertaken here had the benefit of knowing which academic authors had been submitted for the 2001 RAE and, consequently, could additionally quantify their citation counts alone and calculate a correlation between their scores and the RAE rating. The results from this study, like its predecessors, show a significant correlation between citation count and the RAE rating. The robustness of the methods used and consequent results, are consistent with the confidence with which this assessment tool has been used here and in the past. The process is not, however, perfect and it has its limitations.

Overall, the citation study ranked the universities in line with the RAE rating. However, there were anomalies. Birkbeck College achieved a 4 rating whilst submitting just one member of staff for assessment and recording only six citations. Trinity College Carmarthen, likewise recorded six citations with two staff and retained its 3b rating. None of the authors published in the ISI's indexed journals. Confusingly, Carmarthen declined to submit seven other staff for assessment, one of which (M. Patton) would have tripled its citation score. Birkbeck with its average of six citations was correctly placed by the citation study at a borderline 4/5 rating but, on total count, should have been placed with Carmarthen. At the other extreme, Sheffield lost its 5* status whilst still having a very high average citation count. Belfast with three times the average citation rate of its nearest rival and top of the table, did not gain 5* status but retained its 5 rating. Belfast's very high citation rate is attributable to P J Reimer's exceptional score of 565 citations. These outliers at the

extremes of the rating scale, where citation counts are very high or low, is where the RAE panel had to exercise difficult judgements [3].

The RAE Process and Publishing

This study was supplemented with an interview with Professor Barker of Leicester University's School of Archaeology and Ancient History, a member of the archaeology RAE panel. In it, he addressed a number of questions regarding the effect of the RAE on individuals, on departments and on working methods [3].

It would seem that there is a link between achieving 5 or 5* status and publishing in high impact journals, i.e., those identified for indexing in the AHCI. Table 11 demonstrates that the higher the number of AHCI indexed journals you publish in, then the higher your RAE rating and citation score. This is most definitely **not** a cause and effect relationship and University managers who try to make their academic staff only publish in high impact journals are being astonishingly naïve. Important papers tend to be submitted to prestigious journals. Prestigious journals are indexed in the AHCI. Important papers tend to be submitted to the RAE and tend to lead to higher RAE ratings. Professor Barker [3] was at pains to point out that where the article or book was published did not affect the panel's judgement regarding its quality. The threshold to obtain publication in these journals is that much higher than many others and so, by default, the quality of submissions will need to be higher, thus promoting better quality articles and hence a better RAE assessment.

Conclusions and Recommendations

The citation study undertaken here demonstrates that a high correlation exists between the 2001 RAE scores obtained by archaeology departments and a citation count of the academics in those departments. This result confirms the evidence from other studies, which also demonstrate the robustness and accuracy of the correlation. *Staff submitted* correlation scores were marginally better than *all staff* correlation scores, but both were highly statistically significantly correlated with RAE scores.. We hope that with the publication of these results, the debate about the robustness and consistency

of the correlation will be laid to rest. It remains, of course, right and proper that the *implications* of the correlation should be debated.

Assessment of the quality of research output from UK universities will continue, but how it should be conducted is open to debate, a debate that is strongly encouraged in the subtitle of (Roberts, 2003), “Issued for consultation”. There can be little doubt that, whatever the changes in the future, the quality of research outputs will be a major component of the metrics adopted. Whilst citation analysis is not a perfect tool, it is recommended that it should be adopted **as the primary procedure for the initial ranking** of university departments. Once complete, the rankings could then be distributed to the panels for consideration. *We do not recommend that citation counting would be the only assessment tool.* For one thing, by definition it only looks backwards and takes no account of future plans. Peer review of the written evidence, consideration of other factors and the careful scrutiny of marginal cases must complement the process. Nonetheless, we believe, despite the well-known suspicion by academics of citation-based measures, that there is a convincing case that citation analysis should form the first part of any future assessment of research quality.

Notes

[1] Part IV in this series: (Holmes & Oppenheim, 2001)

[2] Ahmed, T., Oppenheim, C. and Parker, S., unpublished results.

[3] Personal communication from Professor Barker, Member of the Archaeology RAE Panel in 2001.

References

EN.REFLIST