Pergamon

PII: S0952-8733(97)00021-4

Higher Education Policy, Vol. 10, No. 3/4, pp. 297–313, 1997
© 1997 International Association of Universities
Published by Elsevier Science Ltd. All rights reserved
Printed in Great Britain
0952-8733/97 $17.00+0.00

FORUM

Citation analysis and Northern Ireland: a quality measure?

Stephen B. Johnson*† and Robert D. Osborne*

**Centre for Research on Higher Education, University of Ulster at Jordanstown, Northern Ireland and †Queen's University of Belfast, Belfast BT7 1NN, Northern Ireland*

Research evaluation is used to identify "success" and relate this to funding. Citation analysis is one of many performance indicators but has been largely set aside in the U.K. This paper describes the use of bibliometric data and examines the "parochialism" of Northern Irish research. Papers produced in Northern Ireland between the years 1981 to 1994 and listed with the ISI are used to exemplify the issues. The analysis indicates that some fields are underrepresented in the ISI database. Small research systems can also be significantly influenced by one or two individuals. Publication in a highly visible, Anglo-American, internationally refereed journal will enhance the citation rate. Northern Irish research has a relative lack of international impact, seemingly a function of topics and the journals used. While there is an increasing amount of joint authorship, particularly with the rest of the U.K., relatively little collaboration has taken place with colleagues in the rest of Europe. Citation analysis is an important initial indicator of research impact, useful to establish questions and narrow an overall field of inquiry. © **1997 International Association of Universities**

The evaluation of research is a major facet of public policy in many countries. The reasons for undertaking research evaluation vary but at the core is a concern to try and identify "success" and, increasingly, relate this to funding. Evaluation itself incurs direct and indirect costs which have to be set against the funding level of the research being evaluated. The object of evaluation can be a single piece of research, a programme, or, as is the case in the U.K.'s Research Assessment Exercise (RAE), the total research activity of a country's universities.

The U.K.'s research assessment exercises during the 1980s and 1990s have moved towards qualitative peer assessment and away from volume and/or quantitative measures. Bibliometric, or citation analysis, much discussed in the 1980s as a possible tool for research evaluation in the U.K., has been largely set aside. Other countries have maintained the use of citation analysis. In North America the quantitative, bibliometric approach is one of many performance indicators. The Higher Education Authority (HEA) of the Republic of Ireland, with a relatively underdeveloped national research evaluation system, has also made use of the quantitative, bibliometric approach (see Osborne, 1996).

This paper has two main purposes: (1) to explore the use of bibliometric data to assess research output and quality; and (2) to examine the impact of research conducted within Northern Ireland. The questions were explored by a review of the citation analysis literature and analysis of the papers produced in Northern Ireland between the years 1981 to 1994 recorded in the Scientific Citation Index of the Institute of Scientific Information (ISI). Comparisons of recorded Northern Ireland research output with other countries and trends in co-authorship are made.

The purpose of undertaking this analysis is to examine a number of issues which have emerged concerning the nature of research in Northern Ireland, as well as interrogating the data for what it can tell us about citation analysis and research output in general.

WHY EVALUATION?

The 1987 OECD report on research evaluation noted that the "identification and preservation of 'worth' is one of the principal aims of evaluation" (OECD, 1987, p. 14). Over the past two decades major questions as to the worth of scientific research have been raised. This questioning reflects changes to both science itself and the socio-economic conditions under which research is now conducted.

The reasons given for the increasing concern with the evaluation of research have often been explained in terms of "economic realities". Changes in the nature of science and higher education, and the nature of western economies (e.g. a desire to obtain technological niches), has also fueled the quality debate. Since the 1980s, limited research budgets, the concept of accountability, the expense of modern research and recession, have created a climate for the critical review of research expenditure. In an environment of limited, some would say declining, resources, new research can only be initiated at the expense of other programs as all research must be carried out within fixed expenditure levels (Aitkinson, 1992; OECD, 1987; Whiston and Geiger, 1992).

As a consequence of these economic changes, policy makers have increasingly turned to research quality analysis to provide formative assessment of the worth of projects in order to make decisions about funding. This economic argument only reflects part of the influences on the development of an evaluation culture. There have been previous recessions and limited research budgets.

The 1987 OECD report suggested that the growth of science and research related activities themselves have been the origin of research resource problems. The growth in science has been related to increasing competition for intellectual leadership and, as a consequence, the question of what is the best science arises. In particular, competition between the institutions themselves has become a significant influence on the development of evaluation processes.

The competition and expansion in research activities has been highlighted by the restructuring and growth of higher education within OECD countries. For example, there has been a massive growth in the university sectors of Australia and the U.K. through the upgrading of technical institutes and polytechnics to university status. In the 1970s the U.K. consisted of some 50 universities, today there are 138 institutions. Concurrent with this expansion is general resource decline in government funding for research (e.g. Whiston and Geiger, 1992). As the majority of current funding system reinforce the traditional university model of research and teaching, the new university institutions have increased the competition for a stable, if not declining, pool of funds.

The exploitation of science for economic growth has also been a significant influence on the emphasis on evaluation. Questions about which science, discipline, or expert will likely provide new technologies to fuel success in some technological niche can only be assessed by intensive evaluation.

THE PEER REVIEW SYSTEM

Evaluation procedures are used both to legitimize current fund allocation as well as to identify where limited funds should be directed in order for institutions and society to compete in a changing world economy. Understanding the evaluation process itself is necessary for critics and proponents of research policies. Evaluation criteria can have substantial impact on research and the future flexibility of agencies (Whiston, 1992).

Within much of the western research system there is one dominant method of research evaluation, obtaining a consensus of expert judgments. The most common form used is the peer review technique, for example, the U.K.'s RAEs of 1992 and 1996. There exist three forms of peer review that differ in the information used by experts to reach a sound judgement:

(1) Direct (e.g. an expert panel);
(2) Modified (e.g. a survey of research peers, users etc.); and,
(3) Indirect (e.g. historic peer review, publication measurement).

One of the more influential techniques has been the use of indirect measures of peer review. Indirect measures move the locus of review from a panel of experts to the relevant community of scientists as a whole. The most significant indirect measure is the assessment of scientific production. Dahllof *et al.* (1991) distinguish four types of scientific production:

(1) The number of theses produced within a department or laboratory;
(2) The rewards obtained by researchers on an individual basis;
(3) The measurement of publications; and,
(4) Research contribution to companies and society.

Each have their own strengths and weaknesses. For example, rewards are often given for work over an extended time frame and long since completed. Rewards are few, however. To some extent they reflect a bias towards the opinion of the scientific community rather than the merit of the research being conducted. The measure of publications has often been the first, and occasionally only, assessment of the quality and impact of research.

The published paper or its equivalent represents the fundamental social-process of science, the communication of ideas (Fox, 1983). Publications also represent ways in which individuals gain recognition, competitive research funds and promotion, and provide evidence of institutional excellence and access to selective research funding (e.g. HEFCE, 1996; O'Neill and Sachis, 1994; Ramsden, 1994).

The assessment of publications generally takes two forms, publication counts and citation analysis. Publication counts are of widespread use representing a simple measure of research output (Harris, 1990). There are generally strong correlations between the impact of a piece of research and quantity (Ramsden, 1994), but overall it is hard to evaluate the true significance of any research output. Publication counts are also problematic when the freedom of communication is restricted and for "gray"

literature such as reports and conference proceedings. It has been suggested that publication counts are more use at institutional levels (OECD, 1987).

Citation analysis relies on the behaviour of scientists to "cite earlier publications because the work contained in them is, in some way, relevant to their own" (OECD, 1987, p. 34). It assumes that the number of citations is a reflection of the influence of articles relative to others and, indirectly, assesses the quality of research, a form of market signaling behaviour. The major benefit of performance indicators such as citation analysis is that comparisons across units, whether individual researchers, laboratories, departments or institutions, can be made. Citation analysis also allows the reviewer/policy-maker to translate difficult judgments into numbers and enable researchers to explore the growth and decline of a field of study.

There are three broad assumptions that underpin the use of citation analysis. It assumes that research output is consistently represented in journals, that the number of citations is a legitimate indicator of quality and impact, and that accurate data are available.

PRECONDITIONS TO CITATION ANALYSIS

Cave *et al.* (1988) describe a number of preconditions necessary before citation analysis and general publication analysis are undertaken. The assessor must review the types of publications to be included, assign weightings for different publication types, define the sources of information about publications and, define the unit of assessment (e.g. university, campus, school, research group, individual).

The types of publication available to any analyst is substantial. Analysis is generally restricted to books, refereed journals, significant journals, conferences and reports. The decision about what to count is dependent on decisions about weightings and the nature of the field being assessed. For example, it has been well established that different publishing modes characterise different subject areas with the humanities and social sciences relying more on publication in book form than the sciences (e.g. Ramsden, 1994).

Often the weightings given to material included in an analysis reflect the order books, refereed journals, significant journals, conferences and reports. Refereed journals are more often analysed, however, due to the more comprehensive databases available to the researcher. Cave *et al.* (1988) suggest that there is an implicit weighting system operating even before the author of any citation review begins, for example, one for a recognized journal and zero for omitted ones. As with what to count, the weightings given to the different publication types depend on the nature of the analysis required (its aims and objectives) and the field being examined. For example, citation analysis has been used to map the current state of a field of research as well as assessing scientific productivity. If the field is represented strongly by a certain set of refereed journals then it may be appropriate to assess the quality of the research and the nature of the field by concentrating on these publication types.

There are three important questions that any reviewer of citations must ask themselves when considering what to count and the weightings to assign. These are, whether to include self-citations, the relative weighting of seminal research and the weighting of collaborative research.

For most analysis self-citations are not included. The issue of seminal articles is slightly more complex. Many authors have suggested that seminal articles are not

given due weighting. Such articles are often not directly referred to and in some cases later work by other authors may be easier to read. Moravcsik (as cited in OECD, 1987) argued that concepts soon become untraceable.

The last of these concepts, collaborative research, has become more important in recent years with the expansion of research centres and institutional links. Many databases used for citation analysis list by the senior/primary author. Hunting for junior author(s) is time consuming and as such their productivity may not be recognized. This is even more important when we consider the development of differing authorship practices such as alphabeticizing, random order or rotation.

There exists a number of relevant databases that supply details of citation analysis, though the more commonly used one is the Science Citation Index (SCI) published by the Institute for Scientific Information (ISI). The SCI database provides the reviewer/researcher with a good coverage of scientific journals. In recent years the database has expanded its base of source journals. An important benefit of using the SCI is the provision of a figure for each journal listed detailing the journal's average number of citations. As a consequence an "impact factor" can be determined by comparing the ratio between the observed citation rate for an article with that journal's average, or expected, citation rating.

The major limitation of citation databases such as those produced by ISI, is their incomplete nature. For example, the SCI journal coverage indexes approximately 70% of scholarly journals, with some 80% of these peer reviewed. Books and other forms of publications are rarely noted. This incompleteness represents a significant possibility of undercounting the impact of researchers in fields where journals, and particularly peer reviewed journals, are not the major forms of publication. As a consequence, the publication traditions within and between fields, and across time, can substantially effect the observability of research units in citation databases. For example, Burnhill and Tubby-Hille (1994) noted that books are generally ignored by most databases whereas in the social sciences books have "greater importance in the dissemination of research findings" (p. 140). They noted that for the social science fields surveyed only one-third of the total published research is actually presented in refereed scholarly journals and as a consequence undercounting of research product is very likely.

It is a mistake to assume that all areas within, for example, social sciences are equally effected. Publication traditions can be quite diverse within a broad field. Burnhill and Tubby-Hille's (1994) analysis of the percentage of material published in journals noted that 55% of psychological, 30% of social anthropology and 12% of linguistics publications were to be found in journals.

There is also a restriction in the journals used by many authors. Burnhill and Tubby-Hille (1994) reported that for a total of 1564 serial titles surveyed, 50% of publications were contained in 215 of the serials, 25% of the articles in 51 titles and 10% in 13 titles. As a consequence, it is possible that the exclusion of one or two important publications from a citation analysis will significantly underrepresent the true state of the research impact.

The bias in publication type also influences its exposure in a databases due to the time differences required for different publications to make an impact. Different journals have different time lags in reception, assessment and printing of research. Books, journals and reports all have different production rates depending on the publisher, length and other factors. For example, social science with its bias towards books suggests a lag between the research and impact can be between 5–10 years. As

a consequence, any analysis of citations must be aware of the age of the research group and make allowances for long-term effects.

The dissemination and research actions of those within a field also influence field exposure in citation databases. For example, social science research is poorly represented in many databases due to the "tradition" of low publication rates with high variation. That is, many researchers apparently produce little while some researchers dominate their field. There appears to exist a laissez-faire approach to the dissemination of research within social science.

Citation indices are also biased towards basic fields of research and have limited coverage of more applied fields. Groups representing the very end of epidemiological research can be at a disadvantage.

The databases themselves are also prone to error. Moed *et al.* (1989) reported problems with missing data such that about 10% of their total data was incomplete due to programming or operational errors. In some cases 50% of a research group's citations were missing due to the incompleteness of data from a small number of journals. The lack of coverage and errors in the databases may mean that one or two highly cited articles can be missed from any analysis with some significant journals entirely ignored.

This lack of coverage is even more relevant for research producers outside the Anglo-American English speaking sphere. Current databases favour articles in the predominant language of predominant journals. Currently English language, Anglo-American, basic research and review journals are at an advantage. The visibility and availability, language, type of publication, regional location and applied nature will influence citation ratings (Luukkonen, 1991).

While considering what to count, how to weight and what data source to use, the citation analyst must also be aware of the impact of the unit of assessment, in particular the effects of size and age. For example, the SCI dataset has problems with small units (less than 10 individuals) as their productivity is often lower than larger groups and can consequently be seriously effected by the incompleteness of citation databases. The restricted number of publications within such databases favour larger departments. Luukkonen's (1991) comparison of a large and small research group noted that the larger research group had more articles per group and was cited more frequently. The correlation between the number of articles and citations per article per year was 0.64 for the large group, while the correlation for the small group was not significant ($r = -0.30$). Luukkonen (1991) suggested that there is a threshold of size and organization to produce cited research recorded in commonly used databases.

As suggested earlier the differences in publication traditions can influence the time required for research to come within the public domain, consequently older research units will have more exposure. The age of a research unit can also impact citations with research suggesting that researchers and institutions exhibit growth and decline stages with an uneven number of citations depending on the unit's age. For example Tiler and Boddington (1993) suggest research centers go through three distinct phases: growth; emphasis on policy/application; and, international growth. Such growth can result in citations that plateau in the middle phase. This pattern is not clear as there exists little research examining the development of publications for any unit of assessment.

There are also other questions such as: who to class as a member of an assessment unit; what happens when someone leaves and does the citation analyst credit the current or past assessment unit of the author; and, how are projects whose data is

used by groups outside the direct project team or whose data and results are confidential and of limited circulation assessed? Such questions become especially important when we consider that any analysis can be skewed by one or two very productive individuals.

We have noted a number of questions about the technical aspects of defining a programme of assessing citations. What to count, how to weight, what database to use and what unit of assessment we are interested in are important technical aspects. All of these technical issues can substantially influence the results of any citation analysis. While technical adequacy is of concern, more fundamental are questions about the link between the citing of research and research quality.

CITATION ANALYSIS AND RESEARCH QUALITY

There are a number of important considerations underlying the use of citations as a tool to assess the quality of research. First, citations can not be assumed to be a measure of research quality as individuals cite for many reasons. Second, citations are indicators of past performance and may not reflect the assessment of research impact and quality by other stakeholders. Third, the quality of research is a reflection of the environment of the researcher. Finally, the practice of citation analysis can influence researcher behaviour.

The most important assumption underlying the use of citations to analyse research is the link between citation rates and the impact and quality of research. That is, the amount of citation is assumed to indirectly represent a measure of the quality and impact of research. This assumption is misleading. The number of citations makes no distinction between the reasons for citing. A number by itself can not distinguish between significant and trivial work and assumes "positive quality" even though the use of a piece of research may have been critical, negative or derogatory. It is difficult to differentiate quality, impact or recognition as cause for citation.

Brooks (1985) has suggested that the basis of criticisms against the use of citations to assess quality rests on the belief that not all citations are made for cognitive reasons. To explore this issue Brooks asked 26 U.S. university faculty staff to rate a piece of work they had cited in a recent paper on seven criteria (currency, negative credit, operational information, persuasiveness, positive credit, reader alert and social consensus). Brooks noted that the motivation to cite differed between humanities and science groupings, though persuasiveness, the citer using a paper to substantiate their own view, was rated high in both groups. Brooks concluded that authors are "advocates of their own points of view who utilize previous literature in a calculated attempt to self-justify" (p. 228).

The citation of a paper may also reflect a weak literature review, the citer's relationship to the author(s), a desire to be published and therefore take account of editorial biases, or a desire for prestige by citing previous highly regarded work. There is also evidence to support a "Matthew effect", that is, well regarded researchers or institutions will be cited more often. For example, Cave *et al.* (1988) noted a high correlation between reputation and total publication rankings for departments. Counter to this Clarke (1971) reported that performance indicators such as citations are unaffected by received reputations and further analysis by Cave *et al.* (1988) noted that correlations between per capita publication rankings and reputations are lower than reputation and total publication rankings.

Citations are also indicators of past performance and gauge impact of work in the short term. Many decision-making systems, however, assume that by summative analysis of the past judgments for the future can be made. Resource allocation based on "good" performance may not lead to the developments expected. "Good" institutions generally have little room to improve and thus decision-makers are faced with diminishing returns for their scarce resources.

Citation analysis also ignores the impact of research in areas not directly related to scientific production such as "social progress". Citation analysis represents the research producers interests and not other stakeholders such as parents and employers who may have quite different assessments as to the effect of a field of research (Cave *et al.*, 1988). Finally, the impact of a piece of work is also determined by the extent to which researchers carry out public relations tasks (Fox, 1992).

Whiston (1992) has argued that accountability shapes the research culture from without. The practice of citation analysis can greatly influence the behaviour of those being assessed. A research evaluation system can not safeguard against, and may encourage, strategic behaviours on the part of the appraised. Using publication counts can and will affect the shape and size of publications.

There is also the question concerning any conflicts in the ratings of direct and modified peer review systems with the indirect method of citation analysis. The current state of affairs seems to suggest that the direct and modified peer review systems are more indicative of the quality of research if conflict occurs.

The environment within which research takes place can also greatly influence the research productivity of an individual. Citation analysis, as with other publication analysis, has been criticized as not being able to fully reflect the context in which academic activity takes place. The effect of Individual (e.g. personality, motivation, stamina, intellectual style) and Environmental (e.g. climate, nature, resources) factors on research productivity must be understood by any reviewer. For example, Fox (1992) found that a researcher's publication rates conform to the local context after 3 years. Fox (1992) suggested that the research orientation and activities of colleagues, collaboration, resources, diversity of activities and perceived and experienced work climate, effect research productivity. Researchers require the company of people to exchange, activate and reinforce ideas. Fox (1992) reported a high correlation ($r=0.70$) between the total number of articles in refereed journals and the number of articles published in collaboration. She concluded that collaboration is important as independent research is difficult to initiate, sustain and fund. Co-authored papers are also more likely to be accepted as they have had another editor and consequently less likely to have major flaws in their construction. Intriguingly Fox's data suggested that productivity may not be related to a comfortable environment, researchers in competitive environments may actually produce more. Fox (1992) concluded that "factors such as size of department, number of active researchers, and numbers within subareas become important as they influence the possibilities for communication and exchange" (p. 107).

Arguments for and against the use of citations to assess research quality concern the significance of the act of citation, the representatives of the data and the degree of agreement with other methods. As a consequence, it seems generally agreed that citations can not be directly matched to measures of research quality, though there is a correlation. Citations represent an important initial indicator of research impact and quality (Dahllof *et al.*, 1991). Bibliometric systems such as citation analysis are a partial and imperfect form of information reflecting different facets

of research performance. It requires the use of different indicators that show convergence.

ANALYSIS OF RESEARCH PRODUCTION IN NORTHERN IRELAND

The concept of citation analysis is particularly relevant in examining research productivity in Northern Ireland. Northern Irish research has often been described as parochial and the U.K. RAE exercises have highlighted substantial differences in the performance of different elements in Northern Ireland's small research community.

In a paper setting out a rationale for a special seminar series based on Northern Ireland, the Economic and Social Research Council (ESRC) suggested that ESRC awards to Northern Ireland academics and the projects that were funded tended to be small and "parochial" in that they were primarily concerned with Northern Ireland (see Osborne, 1993). There is no doubt that the charge of parochialism has stuck and this description now often encompasses research beyond the social sciences.

The overall performance of the two Northern Ireland universities in the U.K.'s RAE, especially in 1992, resulted in a patchy, or middling performance (Osborne, 1996). As a part reaction to this, government in Northern Ireland channeled additional research funding to the universities through the Northern Ireland Development Research fund. This funding sought to ensure that research in Northern Ireland achieved the localist agenda of local economic, social and cultural concerns whilst at the same time enhanced the potential performance of the two universities in the national U.K. research evaluation exercises, a strategy with inherent tensions between local concerns and national excellence.

FINDINGS

In order to explore some of the issues relating to the use of bibliometric techniques, data was obtained from the Circa Group (Europe) for Northern Ireland publications listed with ISI over the period 1981–1994.* This data included details of the 50 most cited Northern Ireland Arts, Humanities and Social Science papers, the 100 most frequently cited Northern Ireland papers, publications rates per annum, citation impact, the top 100 most cited Northern Ireland researchers, co-authorship details and citations trends for Northern Ireland by research field.

ASSESSING THE TOTAL NUMBER OF CITATIONS

This section examines the impact of using the total number of citations to describe research productivity and make decisions as to research quality. Deficiencies are noted with regard to comparing categories of research with different publication traditions. Comparisons between fields were made using a production index. Data obtained from CIRCA for Northern Irish research production between the years 1981–1994 were used for the analysis.

Northern Ireland (NI) has seen a substantial increase in research output within all

*The CIRCA group was contracted by the Higher Education Authority to conduct a similar analysis of the Republic of Ireland.

categories of publications covered by the SCI. The NI publication output increased from 986 publications in 1981 to 1750 publications in 1994. The growth was consistent across the two broad categories of Science & Engineering (S/E) and the Arts, Humanities & Social Sciences (AH&SS). Science & Engineering increased from 774 publications in 1981 to 1257 in 1994, while AH&SS increased from 213 to 355. Over the 14 year period an average of 280 papers per year were listed in the SCI within the AH&SS category, while the S/E category averaged 1352 papers per year. The SCI is heavily skewed towards the sciences. 77% of the SCI recorded publications were classified as Science & Engineering while 21% of the total output (3941) for this period was categorised as AH&SS.

The difference between the production of the two categories may be reflective of a substantial difference in the research publication productivity. Owing to the lack of coverage by the SCI of non-journal publications it is more likely that the Arts, Humanities and Social Science relatively low publication volume is reflective of the database and publication traditions (i.e. books rather than papers). In order to make a comparison of between the two categories we can compare the output of later years using the 1981 output for each category as a base production value, a production index (Fig. 1).

Figure 1 indicates that the level of production in both categories of research for the years 1982–1994 was higher than the 1981 level. The graph also indicates a steady growth in the publications produced in Northern Ireland for the two categories. Generally, the level of production for the S/E category was higher than the level of production for the AH&SS category except for the years 1985, 1986 and 1991, while 1994 saw a equality in production. The figure also indicates that the years 1987 to 1989 are of interest as they exhibit a drop in the upward trend of higher levels of production.

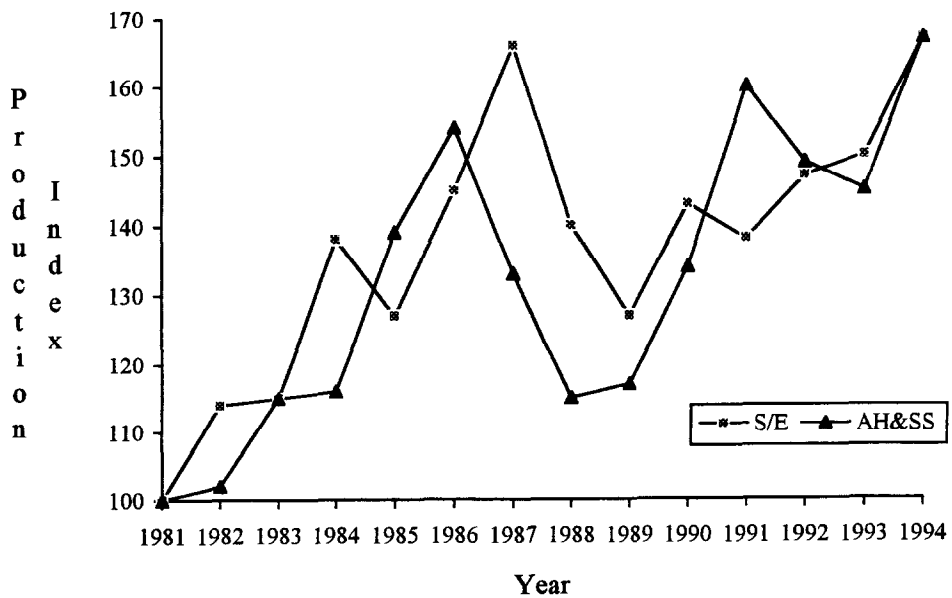

Figure 1: Northern Ireland publication production levels from 1981–1994 for the categories of Science/Engineering (S/E) and Art, Humanities and Social Sciences (AH&SS) using the production for each category in 1981 as a base of 100. Source: CIRCA/ISI (1996).

Overall, the levels of production for the years 1982–1994 for the S/E category were an average of 40% higher than 1981 level, while the levels of publication production for the AH&SS category were on average 34% higher. Both categories are susceptible to large fluctuations in production levels from year to year. For example, S/E increased in production by 21% from 1986 to 1987, and then dropped by 26% from 1987 to 1988. Publications produced in the AH&SS category increased 15% from 1985 to 1986 and decreased 21% from 1986 to 1987.

Care must be taken in being too positive about the overall trend in production growth. As ISI expands its database of listed publications we would expect to see some increase in the number listed per annum. Comparison between national outputs is required in order to establish if this trend of increasing production is similar to or better than other countries.

The growth in Northern Irish publications highlights differences in assessing social science research using citation analysis. The major problem is the low level of social science representation within the citation databases. As a consequence, the total number of publications in different categories become a meaningless concept when comparisons between groups are made. For example, for 1981–1994 none of the top 100 most cited papers for Northern Ireland include any Arts, Humanities and Social Science papers. There is a predominance of medical papers within the most cited papers. If we examine the top 50 most cited AH&SS papers we can also note that the majority of cited papers represent a "medical" bias. For the period 1981–1994, 16 of the 50 most cited AH&SS papers are medically oriented, dealing with areas such as substance abuse, drug side-effects and psychopathologies. Much of the remainder is a mix of experimental and clinical psychology and some economic articles.

This domination of one or two fields in the listings of most cited papers support criticisms about the lack of coverage of non-journal material by databases and reinforces the concept of publication traditions. The dominance of psychological oriented publications in the top 50 most cited AH&SS papers reinforces Burnhill and Tubby-Hille's (1994) finding that psychology is dominated by the use of journals to publish research.

The data presented here indicates that the use of total citations to examine the impact of a publication is problematic. Other approaches to defining the impact of an article or a field may be more indicative of the quality of the research.

OTHER WAYS TO ASSESS IMPACT

The ranking of departments, schools or individuals by the total number citations is generally used within the literature as a direct measure of the quality and impact of research. As noted previously, this biases analysis of social sciences with their low representation in citation databases. The impact of a particular paper can also be assessed by rankings based on two other relationships, the ratio between the observed and expected number of citations, and ranking based on citations per year. The benefit of these two different measures of impact is that they acknowledge the restricted nature of citation databases and different publication patterns across subjects.

The SCI provides a measure of the average number of citations for any particular journal. As a consequence, research impact can be assessed by comparing this expected citation rate with the articles' actual number of citations. This approach has the added advantage in that the journal's visibility to the broader research community is

indirectly accounted for when assessing the impact of research. That is, journals with a wider circulation are likely to provide the means for individuals to obtain higher levels of citations that is not a reflection of the true "worth" of the research but of the journal itself. This is particularly relevant for new researchers who may not be able to publish in the high profile journals with wider circulation, or for researchers who have published in a high profile journal but the article fails to perform. The ratio method assumes that articles with ratios greater than one have made a significant impact in their area that goes beyond the effects of the journal's visibility. Articles with ratios less than one have made limited impact in the field. The drawback is the assumption that the quality and impact of journals are equal. That is, an article in an internationally recognised and refereed journal is equivalent to an article in a more specific, locally refereed journal. It also is limited in that it can not give reasons for citing, such as the refuting of arguments presented in a cited paper.

Analysis of the 50 most cited AH&SS articles provided a range of observed to expected citations ratios of 0.40 to 20.31 with an average ratio of 2.70. 60% of the articles had a ratio of two or lower while 19% of the articles had a ratio less than one, representing articles which were cited less than expected. The ratio of 20.13 was an extreme value for the dataset as the next highest ratio was 6.69. With the extreme ratio article removed the average ratio dropped to 2.33. The data indicates that for AH&SS articles listed in the SCI their observed citations were, on average, approximately twice that expected.

For the 100 most cited articles from Northern Ireland the ratio between observed and expected ranged from 0.54 to 17.81, with an average ratio of 4.50. The ratio of 17.18 was not an extreme value as the next highest ratio was 16.12 and 6% of the articles listed had a ratio greater than 10.00. 4% of the articles had a ratio less than one and 19% of the articles had a ratio of less than two. The analysis indicates that for the 100 most cited articles (mostly medical/scientific papers) the observed citation rate was approximately four times the expected citation rate for a journal.

The ratios for the two sets of most cited articles exhibit differences in the assumed impact of the articles on the wider community. In general the 100 most cited articles were likely to have a citation rate four times the level expected while very few of the articles were cited less than expected. In contrast, the 50 most cited AH&SS articles were rarely cited at levels more than twice the expected citation rate and a substantial percentage were cited less than expected. This difference in citation ratios between the two fields indicates cause for concern about the applicability of Northern Irish AH&SS papers to the wider scientific community.

The use of ratios can significantly effect the ratings of the impact of an article and thus its assumed quality. Providing a measure of total number of citations does not account for the influence of the journal on artificially inflating this figure. Rankings based on ratios of observed to expected citations can produce a different order. For example, there was no significant correlation between the rankings of the 50 most cited AH&SS papers by ratio and by total number of citations. For the rankings of the 100 most cited papers the correlation was 0.24 ($P<0.05$), significant but low. That is, for the AH&SS papers the rankings of "impact" were significantly different depending on the system used to rank the articles. For the 100 most cited articles the rankings based on these two approaches were not significantly different but there was still discrepancies between the two systems.

The use of ratios rather than citation counts to assess the impact of an article can alter the nature of the relationship between AH&SS and the S/E articles that dominate

the most cited lists. As noted previously, none of the 50 most cited AH&SS articles appeared in the list of the 100 most cited Northern Ireland papers. A different relationship was noted when the two lists were merged and ranked on the ratio between expected and actual citations. For the combined ratio list of 150 articles the top 50 articles included eight from the AH&SS database with the highest ratio an AH&SS article. The use of ratios, implying a measure of the impact of the research on the wider research community, can have serious effects on the rating of publications and research groups.

The comparison of the two sets of most cited articles indicates that the scientific based articles (generally S/E articles) were making a greater impact in their field than the AH&SS articles. In particular, the substantial percentage of "most cited" AH&SS articles with a ratio less than one indicates concern for the applicability of this research to the wider scientific community. The use of ratios can greatly influence the perceived impact and consequent ratings on the quality of research. It provides a substantial counterpoint to the dominance of S/E based research in citation analysis.

The second method to assess the impact of an article is to explore the citation rate per year since publication. This measure takes into account the length of time an article has been within the public domain. This procedure assumes that articles with higher rates of citations per year of "public domainess" represent a more significant article as "better" articles should have higher rates per year due to their more significant long-term impact.

The 50 most cited AH&SS articles' citation per year rate ranged from 1 to 15, with an average of 2.84 citations per year, though this was substantially influenced by one paper printed in 1993 which had 15 citations in the first year. The median rate was 2.25 citations per year and 95% of the articles were cited between 1 to 5 times per year. With the atypical case removed the average citation rate for the dataset dropped to 2.50 citations per year. For the 100 most cited papers the citation rate per year ranged from 4.31 to 78, with a mode of 5 and a average of 10.49 citations per year. 95% of the articles were cited between 1 and 20 times per year. The removal of an article that rated 78 citations per year reduced the mean rate to 9.85 citations per year.

The descriptive data suggests that the 100 most cited articles were more likely to have a better rate of citations over their lifetime than the 50 most cited AH&SS articles. The most cited AH&SS articles had a citation per year rate that was, on average, some 20% of the level of the 100 most cited articles. There was an extremely low rate of citing per year for AH&SS articles. This indicates differences in the use of such articles by the wider science community and reinforces differences between fields in their citation practices and also a possibility of a lack of exposure of Northern Irish research in the international research community.

As with ranking based in the ratio between observed and expected citation rates, the use of the "citation per year" figure to rank articles can have a significant effect on the listings. For the rankings of the 50 most cited AH&SS papers there was no significant relationship between the rankings by citation per year when compared to the rankings by ratio. As citations per year represents a transformation of total number no analysis was conducted.

For the 100 most cited papers the correlation between rankings by citations per year and ratio was significant but low ($r=0.18$, $P<0.05$). As with the ratio ranking when ranking the 150 articles by citations per year, the rankings were significantly altered with the top 50 articles including two from the AH&SS dataset.

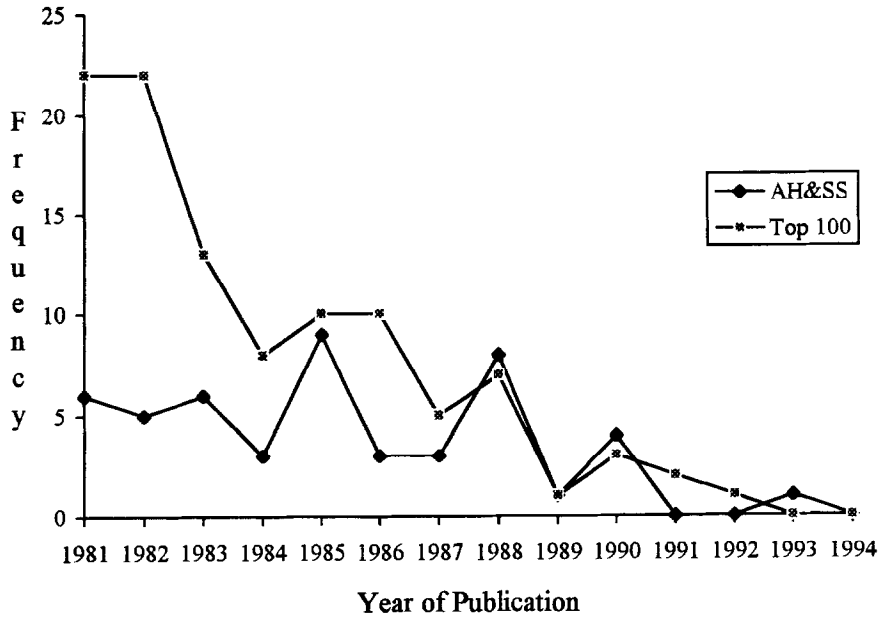

Figure 2: Frequency for year of publication for 100 most cited papers and 50 most cited Arts, Humanities and Social Science publications from Northern Ireland (1981–1994). Source: CIRCA/ISI (1996).

As with ratio measures the use of citations per year has potential to significantly influence the assumptions about the quality of research. This is particularly relevant when considering AH&SS articles with their low level of exposure to citation databases and their differing citation practices. Significantly, the most cited AH&SS papers are relatively poor cousins of the 100 most cited papers having low levels no matter which assessment of impact is used. The use of these two approaches to assessing impact with the different results suggest that the assessment of "impact" is not straightforward. Consideration of the inclusion of variables such as the length of time since publication and type of journal in the assessment of impact is required.

THE LIFETIME OF AN ARTICLE

Groups such as ISI and the HEFCE have assumed that articles generally make their most significant impact in the period 3–5 years after publication. Figure 2 presents the frequency of the date of publication for the 100 most cited papers and the 50 most cited AH&SS publications.

This figure indicates that very few of the most cited publications were less than 5 years old, the standard for assessing the impact of an article. For the majority of the 100 most cited articles they were at least 10 years old, i.e. published before 1984. In comparison, the majority of the 50 most cited AH&SS papers were at least 6 years old, i.e. published before 1988. This difference is suggestive of the reduced currency of social science research over the long-term while more scientific/engineering oriented work maintains currency. The difference in age of articles for the two lists suggests that more defined subject fields (e.g. Fine Arts, History) may have differing relationships between citations and time since publication. As a consequence, any assessment of a

field must be aware of the nature of the knowledge developed and how this knowledge retains "currency" over time.

NORTHERN IRISH RESEARCH: THE PAROCHIALISM ACCUSATION

It has been asserted by the ESRC that Northern Irish social science research is "parochial" and this view has become a more general description of all research in Northern Ireland. This description of Northern Irish research as parochial is made from a particular social and geographic location. It has been argued that much of the social science research in Northern Ireland which is related to the "conflict", while of significant interest to those concerned with examining divided societies in conflict and their management, is of little interest to those in the rest of the U.K. (Osborne, 1996). While the SCI does not allow the total confirmation or refutation of this description it does allow exploration of some aspects of the charge.

One of the major indicators that can be used to explore how much of an impact Northern Irish papers make on the world scene is their relative number of citations compared to the world average. Data from the SCI indicates that Northern Irish research is cited less than expected in comparison to the world average. Since 1981 the rate of citations per paper for all Northern Irish research has been, on average, 10.7% lower than expected. The trend since 1984 has been a decrease in this difference in the citation ratio suggesting that Northern Irish papers are becoming more acceptable to the broader research community. The SCI data also indicates that this difference in expected and actual citation rates is consistent across all fields of study, except for Arts and Humanities, though the number of papers recorded and analysed in this category are very small. Whether this lower than expected performance reflects parochialism in content or a tradition of not aiming for publication in international journals listed in the SCI is open to debate.

There has also been a trend over the years 1981–1994 for an increase in co-authorship between institutions within Northern Ireland, and between Northern Ireland institutions and the rest of the world. Over the period 1981–1994, there were a total of 5812 co-authored papers, representing 31% of the all the Northern Irish papers published. Figure 3 details the percentage of all the papers for the years 1981

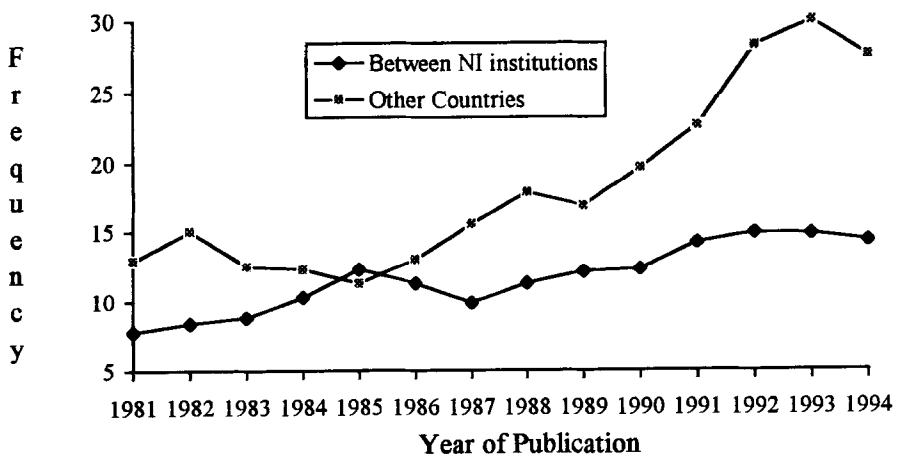

Figure 3: Percentage of total Northern Ireland papers that were co-authored within Northern Ireland and with another country for 1981 to 1994. Source: CIRCA/ISI (1996).

to 1994 that were co-authored between Northern Irish institutions and between Northern Irish institutions and other institutions around the world.

The percentage of co-authored papers between Northern Irish institutions increased from 7.8% of all papers in 1981 to 14.3% in 1994. The last 4 years of the survey period indicates a relative stability in co-authorship trends between Northern Ireland institutions. Co-authorship trends between the Northern Irish institutes appear stable and relatively low. There does not appear to be a trend to develop greater research relationships between the few research institutions that exist.

Co-authorship between Northern Ireland institutions and other countries has, in contrast, seen a major increase. For all years but 1985, the level of international co-authorship has been at a level far higher than between Northern Ireland institutions. Analysis of the countries with which Northern Irish institutes co-authored over this period indicates a dominance of work with co-authors in England. 30% of the co-authored papers were with English institutions. In total, 41% of the "international" co-authored papers were with U.K. institutions. Another 31 countries are recorded in the co-authored lists for 1981–1994. Seventeen of these represent other European countries such as Ireland, Spain and Germany, while seven are members of the Commonwealth (not including the member states of the U.K.). The country with the second highest co-authorship trend (13% of papers) is the U.S., indicating strong cultural and political links with U.S. institutions.

Care must be taken in drawing too strong a conclusion about the nature of co-authorship trends due to the relatively small volume of co-authored papers produced over this 14 year period. For a number of countries there were only 14 co-authored papers, an average of one per year. What the data does suggest is that there are some significant trends in co-authorship such as the increasing number of co-authored papers and the diversity of the countries with which Northern Irish researchers have contact with. Comparisons with other countries are required in order to make firmer diagnostics of the co-authorship and its relationship to the assumed parochialism of Northern Irish research.

CONCLUSION

In drawing conclusions from this paper it is important to reiterate some of the limitations of the data and of the use of citation analysis for evaluating research. Foremost of the limitations is the concentration on journals in the SCI. As a consequence, fields such as the Arts/Humanities and Social Sciences, where publications are often in book form, are underrepresented in the systems commonly used to assess research. Those papers that are represented from this area are dominated by medically oriented topics. The second concern is the size of the research establishment. The Northern Irish data indicates that in small systems one or two individuals can significantly effect a citation analysis. In addition, while citation rates will generally match the average citation rate for a journal, publication in a highly visible, Anglo-American based, internationally refereed journal will positively enhance the citation rate. Access to such journals may not only reflect the quality of the research but also editorial policy. Nevertheless, citation analysis is useful to establish questions and narrow an overall field of inquiry. While it is unlikely that citation analysis will ever fulfill the criteria to represent a good and acceptable performance indicator (e.g. Beecher and Kogan, 1992), not the least because of the costs of collection and

problems in using incomplete databases, citation analysis remains an important initial indicator of research impact. The use of other techniques to assess research quality is, however, necessary if research in different fields within differing traditions is to be suitably acknowledged.

In terms of the extent to which we are able to assess the assertion that research in Northern Ireland is parochial, the jury is still out. Research which is oriented towards the particular social/political context of Northern Ireland will always be limited in its international appeal. Northern Irish scientific/medical research is not, however, of limited appeal. Thus the relative lack of impact is probably attributable to the type of research or the journal location of the publication. The increasing amount of joint authorship with authors from outside Northern Ireland suggests a trend away from a parochial orientation. Links with the rest of the U.K. seem to be strong but it is noticeable that relatively little collaboration has taken place with colleagues in the Republic of Ireland. An agreed political settlement may provide an impetus for further joint work. Noticeably, Northern Ireland researchers have relatively little collaboration with the rest of Europe (it is at a much lower level than the Republic of Ireland) an area of collaboration it would be wise to enhance.

REFERENCES

Aitkinson, H. (1992) Issues in funding research. In *Research and Higher Education: The United Kingdom and the United States*, eds T. G. Whiston and R. L. Geiger, SRHE, London.

Beecher, T. and Kogan, M. (1992) *Process and Structure in Higher Education*, 2nd edn. Routledge, London.

Brooks, T. (1985) Private acts and public objects: an investigation of citer motivations. *Journal of the American Society of Information Science* **36,** 223–229.

Burnhill, P. M. and Tubby-Hille, M. E. (1994) On measuring the relation between social science activity and research publication. *Research Evaluation* **4** 3, 130–152.

Cave, M., Hanney, S., Kogan, M. and Trevett, G. (1988) *The Use of Performance Indicators in Higher Education*. Kingsley, London.

Clarke, K. E. (1971) *America's Psychologists: A Survey of a Growing Profession*. APA, Washington, D. C.

Dahllof, U., Harris, J., Shattock, M., Staropoli, A. and in't Veld, R. (1991) *Dimensions of Evaluation: Report of the IMHE Study Group on Evaluation in Higher Education*. Kingsley, London.

Fox, M. F. (1983) Publication productivity among scientists in modern industrial research laboratories. *Social Studies of Science* **13,** 285–305.

Fox, M. F. (1992) Research productivity and the environmental context. In *Research and Higher Education: The United Kingdom and the United States*, eds T. G. Whiston and R. L. Geiger, SRHE, London.

Harris, G. T. (1990) Research output in Australian university economics departments: an update for 1984–88. *Australian Economic Papers* **29** 55, 249–259.

HEFCE (Higher Education Funding Council of England), *Funds for Research*, Online at http://back.niss.ac.uk/education/hefce/funding/researchfunds.html.

Long, J. S. (1978) Productivity and academic position in the scientific career. *American Sociological Review* **43,** 889–908.

Luukkonen, T. (1991) Citation indicators and peer review: their time-scales, criteria of evaluation, and biases. *Research Evaluation* **1,** 21–30.

Moed, H. F., Burger, W. J. M., Frankfort, J. G. and van Raan, A. F. G. (1989) The use of bibliometric data as tools for university research. In *Evaluating Higher Education*, ed. M. Kogan. Kingsley, London.

O'Neill, G. P. and Sachis, P. N. (1994) The importance of refereed publications in tenure and promotions decisions: a Canadian study. *Higher Education* **28,** 427–435.

OECD (1987) *Evaluation of Research*. OECD, Paris.

Osborne, R. D. (1993) Research policy: a Northern Ireland perspective. *Government and Policy* **11,** 465–477.

Osborne, R. D. (1996) *Higher Education in Ireland: North and South*. Kingsley, London.

Ramsden, P. (1994) Describing and explaining research productivity. *Higher Education* **28,** 207–226.

Tiler, C. and Boddington, A. (1993) Outputs, structure and process in the evaluation of social science research centres. *Research Evaluation* **3,** 107–116.

Whiston, T. G. and Geiger, R. L. (1992) Eds., *Research and Higher Education: The United Kingdom and the United States*. SRHE, London.