DISCUSSION PAPER SERIES

IZA DP No. 2070

An Examination of the Reliability of Prestigious Scholarly Journals: Evidence and Implications for Decision-makers

Andrew J. Oswald

April 2006

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

An Examination of the Reliability of Prestigious Scholarly Journals: Evidence and Implications for Decision-makers

Andrew J. Oswald
University of Warwick
and IZA Bonn

Discussion Paper No. 2070
April 2006

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
Email: iza@iza.org

IZA Discussion Paper No. 2070
April 2006

ABSTRACT

An Examination of the Reliability of Prestigious Scholarly Journals: Evidence and Implications for Decision-makers*

In universities all over the world, hiring and promotion committees regularly hear the argument: "this is important work because it is about to appear in prestigious journal X". Moreover, those who allocate levels of research funding, such as in the multi-billion pound Research Assessment Exercise in UK universities, often come under pressure to assess research quality in a mechanical way by using journal prestige ratings. The results in this paper suggest that such tendencies are dangerous. It uses total citations over a quarter of a century as the criterion. The paper finds that it is far better to publish the best article in an issue of a medium-quality journal like the Oxford Bulletin of Economics and Statistics than to publish the worst article (or often the worst 4 articles) in an issue of a top journal like the American Economic Review. Implications are discussed.

JEL Classification: A11, O3

Keywords: citations, research productivity, economics journals, Research Assessment Exercise

Corresponding author:

Andrew Oswald
Department of Economics
University of Warwick
Coventry CV4 7AL
United Kingdom
Email: andrew.oswald@warwick.ac.uk

* For helpful suggestions, I thank Danny Blanchflower, Gordon D.A. Brown, David De Meza, Amanda Goodall, Ian Walker, Ken Wallis, and Michael Waterson.

An Examination of the Reliability of Prestigious Scholarly Journals: Evidence and Implications for Decision-makers

"The results of the exercise… will be expressed as quality profiles of research in each department submitted to the RAE. They will determine the annual distribution of more than £8 billion for research in UK higher education institutions over a six-year period from 2009." Research Assessment Exercise documentation. www.rae.ac.uk

1. Introduction

The United Kingdom government's forthcoming Research Assessment Exercise will determine how much money goes to each department in more than 100 UK universities. To do this, a panel of experts will assess the quality of every department in every university. Each scholarly article and book will be given by the appropriate panel a quality rating of 4* down to 1*, where 4* corresponds to the highest world-class standard, and 1* corresponds only to a national standard of excellence. On these assessments will turn billions of pounds.

Partly because of the size of the undertaking, there will be pressure -- if only covertly -- on members of these RAE panels to use journal labels (X is a 4* journal, Y a 2* journal, and so on) in a heavily mechanical way to decide on the quality of articles. Rumours of this, and guesstimates of the key list of journals, are currently circulating.

It might seem natural that the panels should behave in this way. An obvious argument could go: these papers have already been anonymously refereed, so the quality of a journal paper will be accurately captured by the prestige of the journal in which it has been published. Thanks to sources such as the ISI Web of Science database, journal standing can be judged fairly objectively, by, for example, 'impact factors'.

In a similar vein, in universities all over the world, promotion committees routinely hear the argument: "this is important work because it is about to

appear in prestigious journal X". But how persuasive is such an argument? There appears to have been little research directed at that question.

As in most areas of life, prestige ratings in academia have their uses, and it is unlikely that any scholar would argue that labels are meaningless. Yet that does not mean that journal names are genuinely a sufficient statistic for quality.

This paper explores the reliability of prestige labels. It collects data on the accumulated lifetime citations to papers published a quarter of a century ago. The data come from issues of the American Economic Review, Econometrica, Journal of Public Economics, Economic Journal, Journal of Industrial Economics, and the Oxford Bulletin of Economics and Statistics. These data show the expected ranking. However, and more interestingly, they also reveal that the best article in a good-to-medium quality journal routinely goes on to have 10 times the citations impact of the 'poor' articles published in issues of more famous journals. This fact is probably not known to most of the people who sit on funding councils, or even to most economists.

2. Data collection and analysis

Assume that after some decades the quality of a journal article is approximately known. Perhaps the most usual measure is that of impact as captured by the total citations the article has received (that is, the number of times the article has been quoted in later researchers' bibliographies).

There is a considerable line of work that uses citations to assess intellectual output and productivity, and it has long been known that professorial salaries are correlated with researchers' lifetime citations, and that these citation counts are a good predictor of Nobel and other prizes. See, for example, Bayers (2005), Moore et al (1998), Thursby (2000), Toutkoushian (1994), Laband (1990), and Van Raan (1998). As is also well-known, citations are a noisy signal of quality -- survey articles tend to garner citations more easily than regular papers, there may be some pro US bias in citations, citation

numbers are more open to manipulation than are publications figures, for some individuals self-citations can cause problems, and so on -- but a common view is that citations are probably the most persuasive single measure of scholarly productivity.

For this paper, a selection of economics journals was taken from the year 1981 (namely, a quarter of a century earlier, to allow a long lag for the 'true' quality of a journal paper to be revealed). The winter issue of the year was examined for the American Economic Review, Econometrica, the Journal of Public Economics, the Economic Journal, the Journal of Industrial Economics, and the Oxford Bulletin of Economics and Statistics.

The AER and Econometrica are routinely viewed as two of the most prestigious journals in economics; in rankings they often appear near or at number 1 and number 2 out of approximately 200 economics journals. The Journal of Public Economics and the Economic Journal are usually viewed as good journals -- routinely in the world's top-20. The Journal of Industrial Economics and the Oxford Bulletin of Economics and Statistics are typically put a little lower again in prestige. They often appear around number 40-50 in journal rankings. At the time of writing, for example, the Web of Science total-citations rankings in the Economics category put the AER and Econometrica at #1 and #2, the EJ at #9, Journal of Public Economics at #16, Journal of Industrial Economics at #47, and Oxford Bulletin of Economics and Statistics at #51.

Data on total lifetime citations were collected on each article. The raw data are summarized in the appendix. Table 1 lays out a summary of the data. As is known, the skewness of citation numbers implies that the mean values lie far above the median values. A small group of papers accounts for the majority of citations.

The remarkable variation in the number of times these journals' approximately one hundred articles have been cited by other researchers is clear from the raw data. The single most-cited paper is the famous theoretical analysis of

trade unions by Ian McDonald and Robert Solow. Published in the American Economic Review, this paper has garnered 401 cites to date. The next most influential paper is the Hausman Taylor econometric specification test published in Econometrica; it has been cited 355 times.

However, many of these papers attracted very small numbers of cites. For instance, over a quarter of a century 15 of the articles have been cited either zero times or on only one occasion. Judged from the perspective of the time elapsed, it might be argued that these articles' contribution to intellectual output has been and probably will continue to be zero. In a sense, their publication might now be viewed as having been an error (with the benefit of hindsight, needless to say).

The mean lifetime cites across these six journals follow the broad pattern that might be expected. The prestige labels are, in a sense, correct: AER 68 cites; Econometrica 63 cites; JPubEcon 22; EJ 30; JIE 9; OBES 7. The top journals thus dominate. Similarly, median lifetime cites are: AER 23 cites; Econometrica 22 cites; JPubEcon 9; EJ 11; JIE 3; OBES 2.

However, the variation of true quality -- as measured by cites -- is strikingly large. Because of this high variance, the less highly-cited articles in the top journals are easily bettered by good articles in less prestigious outlets. For instance, the 4th most-cited article in the entire sample is that by Mansfield et al, which appeared in the Economic Journal, and not in one of the top-two journals. As another example, in the American Economic Review, which is perhaps the most famous journal in the discipline, in its winter issue in 1981 more than one third of the issue's articles had after a quarter of a century each been cited fewer than 20 times. The very best papers in the other lower quality journals had by then garnered far more mentions in others' bibliographies -- respectively 88 cites (Sandmo in the Journal of Public Economics), 199 cites (Mansfield et al in the EJ), 43 cites (Teece in the Journal of Industrial Economics), and 50 cites (Sen in the OBES).

Consider, as a benchmark, the median number of cites. In the two top journals here, it is approximately 22. A natural question is then: how many of the articles published in the other four journals turned out to exceed that level? These 'should', in some sense, have appeared in the top journals. The answer is approximately 16% of the articles. In the Journal of Public Economics, 1 out of 6 does. In the EJ, 4 out of 15 do. In the Journal of Industrial Economics, 2 articles out of 17 do. In the OBES, 1 out of 11 does.

One way to make this point more strikingly is to take the mean value of cites among the 4 least-cited articles in each of the six journals. As shown in Table 1, those totals are respectively 6 cites; 5 cites; 23 cites; 3 cites; 4 cites; and 1 cite. Compared to the best article published in the lesser journals, these are of the order of one-tenth as cited.

Ex post, therefore, labels cannot be relied upon to be free of significant error. It appears that the journal system often allocates high-quality papers into medium-quality journals, and vice versa.

Although the implication of these data is that labels work too imperfectly to be taken as a sufficient statistic for the quality of an article, this does not automatically mean that peer reviewers can ex ante improve upon the journal labels. Perhaps the label is the best that can be done without waiting for 25 years?

Nevertheless, simple evidence against such a view comes out of the raw data. There are signs that the journal editors had an idea which would be the best papers in that issue of their journal. In the way they assigned the order of publication, those editors turned out, ex post, to have what now, in 2006, looks like prior insight. This can be seen informally by looking at the raw data. If we regress total cites, y, on publication-order in the journal, x, (that is whether the paper was first, second, third…eighteenth), we get a more formal sense for the pattern. [Notes and Comments, it should perhaps be emphasised, were omitted from the data; the criterion was whether the papers had these words in their titles]. Summarizing as regression lines:

Econometrica Cites = 133.14 – 7.36Order
AER Cites = 119.43 – 5.41Order
EJ Cites = 66.68 – 4.57Order
JPubEcon Cites = 58.93 – 10.60Order
JIndEcon Cites = 13.15 – 0.44Order
OBES Cites = 19.42 – 2.05Order

Individually, the sample sizes here are too small to give well-determined results (the six results vary in statistical significance from approximately the 5% significance level to approximately the 30% level), but as a group they paint a more persuasive picture.

What editors know, and exactly how, seems worth exploring in future research, because of the importance of peer review in the allocation of research funding in western society. It is possible that it can be conveyed to the experts who sit on funding bodies.

3. Objections and counter-objections

Some natural concerns deserve consideration.

One objection is that the data set used here is small. This is clearly true, but perusal of the Social Science Citations Index shows that these characteristics are found repeatedly. The same general patterns occur, for example, in the winter American Economic Review issues for the later years of 1982, 1983, 1984 and 1985. Looking at the 'worst' 4 articles in each issue, none of these articles reaches 10 citations after a quarter of a century. While it might be useful for other reasons to extend the sample size, the paper's main findings will not change.

A second objection is that citations -- some will say -- should be weighted by the importance of the journal doing the citing. Opinions differ on the case for this. One view is that it is only in the short run that a citation in a top journal

matters more -- because in the long run the issue is the stock of intellectual influence across the whole of the subject as measured by total cites in the year the article entirely ceases to be mentioned. For the purposes of the present paper, however, the key point seems to be that the broad ideas are not going to be altered by weighting the cites totals, because the papers in AER and Econometrica garnering very few cites are not -- it is straightforwardly checked -- getting them disproportionately in the top journals.

Third, it could be argued that self-citations are better removed for the data sample. On balance, however, it seems appropriate not to do so here. It does not alter the conclusions of the paper (because self-cites are insignificant for important articles' total cites), and, for some of these highly influential researchers, there seems a logical case for leaving in own-mentions to those authors' important earlier papers.

4. Conclusions

It is dangerous to argue that publication in famous journal X means that a paper is more important than one published in medium-quality journal Y. This does not mean that young researchers ought to ignore top journals, nor that government research funders should. Nevertheless, the publication system routinely pushes high-quality papers into medium-quality journals, and vice versa.

Unless funding bodies -- including the panel members who are to act in the £8 billion pound Research Assessment Exercise in the UK -- and researchers realize this fact, they may make bad allocative choices. It is likely that some senior scholars already understand the general point made in this paper, but young researchers and funding agencies may not.

According to the data, scholarly articles that appear in better journals do go on, as might be expected, to be more highly cited. In that sense, the journal label carries valuable information. However, there is a noticeably imperfect

match between the quality of the journal and the lifetime cites of the individual articles. Approximately 16% of articles in the four lesser journals studied here ended the period with more citations than the median cites of an article in the two elite journals, the AER or Econometrica. To make the point in a different way, if the criterion is intellectual impact measured by citations, in this sample it was far better to publish the top article in an issue of the Oxford Bulletin of Economics and Statistics than to publish all four of the bottom-4 papers in an issue of the American Economic Review.

If peer reviewers, of the kind who sit on RAE panels, have expert information that allows them to judge quality, then the results in this paper suggest that there is a case for them to do so. They should not rely simply on mechanical rules based on journal labels. It might be objected that perhaps peer reviewers have no extra information that would allow them to rank journal papers (beyond the prestige of the journal itself). This possibility deserves to be taken seriously and needs further study. Nevertheless, one counter argument is to look at the citation levels of the journal papers by order of where the paper appeared in the issue of the journal. The early-position papers, such as the Cooley-Leroy and Rosen papers in the 1981 AER, are more highly cited than articles lower down the order of appearance. This suggests that editors had some ability to forecast which would turn out, 25 years later, to be the best papers. Reviewers of the sort who sit on RAE panels may be able to do the same.

Because these issues arise every day -- in countless funding, hiring and promotion meetings from Hemel Hempstead to Hong Kong -- more research on the reliability of prestige labels would be valuable.

Raw data on the total cites to each 1981 article (in the order they appeared in the journal issue)

American Economic Review
Cooley-Leroy 118
Rosen 123
Kohn 23
Howe-Roemer 8
McDonald-Solow 401
Hendershott 16
Spulber 19
Bresnahan 156
Azariadis 16
Jonung 23
Startz 3
Darity 3
Caves et al 147
Akerlof-Main 45
Walker 0
Mussa 70
Conybeare 0
Boland 53

Econometrica
Malinvaud 28
Hausman-Taylor 355
Mundlak-Yahav 1
Nickell 258
Geweke 40
Godfrey 21
Anderson 17
Bourguignon 11
Harris-Raviv 97
Edlefsen 21
Deaton-Muellbauer 32
Pollak-Wales 142
Balk 1
Helpman 7
King 23
Nakamura-Nakamura 80
Bell 2
Rob 1

Journal of Public Economics
Sandmo 88
Courant-Rubinfeld 9
Hey-Mavromaras 9
Weymark 5
Bennett 0
Berglas 20

Economic Journal
Harris-Purvis 12
Malcomson 44
Bingswanger 77
Dervis et al 7
Mansfield et al 199
Hughes-McCormick 54
Spencer 4
Von Ungernsternburg 15
Skott 0
Chiplin 6
Hughes et al 0
Shah-Desai 11
Masuda-Newman 3

Formby et al 20
Shea 0

Journal of Industrial Economics
Williams-Laumas 13
Lynn 2
Aaranovitch-Sawyer 3
Levine-Aaronovitch 7
Teece 43
Thompson 21
Dries 2
Feinberg 2
White 3
Smith 23
Likierman 0
Hirschey-Pappas 2
Highton-Webb 3
Lamm 15
Bartlett 6
Baye 3
Link-Long 7

Oxford Bulletin of Economics and Statistics
Sen 50
Banerjee 8
Boltho 0
Stromback 0
Winters 0
Mayhew-Rosewell 5
Lye-Silbertson 1
Metwally-Tamaschke 2
Tsegaye 0
Brundell et al 9
King 3

References

Bayers, N.K. (2005) Using ISI Data in the Analysis of German National and Institutional Research Output, *Scientometrics*, 62, 155-163.

Laband, D. (1990) Is There Value-Added from the Review Process in Economics? Preliminary Evidence from Authors, *Quarterly Journal of Economics*, 105, 341-352.

Moore, W.J., Newman, R.J. and Turnbull, G.K. (1998) Do Academic Salaries Decline with Seniority?, *Journal of Labor Economics*, 16, 352-366.

Thursby, J.G. (2000) What Do We Say about Ourselves and What Does it Mean? Yet Another Look at Economics Department Research, *Journal of Economic Literature*, 38, 383-404.

Toutkoushian, R.K. (1994) Using Citations to Measure Sex-Discrimination in Faculty Salaries, *Review of Higher Education*, 18, 61-82.

Van Raan, A.F.J. (1998) Assessing the Social Sciences: The Use of Advanced Bibliometric Methods as a Necessary Complement to Peer Review, *Research Evaluation*, 7, 2-6.

Table 1

Data on the Accumulated Lifetime Citations to Articles Published a Quarter of a Century Ago in Six Economic Journals

Data on articles published in 1981

	American Economic Review	**Econometrica**	**Journal of Public Economics**	**Economic Journal**	**Journal of Industrial Economics**	**Oxford Bulletin of Economics and Statistics**
Mean cites per article in that issue	68	63	22	30	9	7
Median cites per article in that issue	23	22	9	11	3	2
Combined cites to the 4 least-cited articles in that issue	6	5	23	3	4	1
Cites to the single most-cited article in that issue	401	355	88	199	43	50

Notes: These are taken, for each journal, from the winter issue of the year 1981. The data measure the number of times over the ensuing 25 years that the articles were cited by others. The source is the Web of Science's Social Sciences Citations Index, in late-March 2006. The data include short papers, partly because some of them are highly cited, and partly because it was not possible to draw a dividing line between those and full papers, but exclude articles denoted Notes, Book Reviews and Comments (where it was possible to assign these categories unambiguously).