

Prestige Labels

Andrew J. Oswald
Department of Economics
University of Warwick, UK

November 2006

Acknowledgments: This article draws partly upon results described in Oswald (2006). For helpful discussions on these issues, I thank Danny Blanchflower, Gordon D.A. Brown, David De Meza, Amanda Goodall, David Laband, Steve Nickell, Mark Stewart, Radu Vranceanu, Ian Walker, Ken Wallis, and Michael Waterson. My thoughts have been heavily influenced by the ideas of, and electronic conversations with, David N. Laband. Michael Waterson was the first to suggest to me that order of articles within a journal might be correlated with their citations.

Introduction

Journal prestige labels -- this is an A journal while that is a B journal -- are deeply unreliable as a guide to the quality of individual articles. I think this fact is not as widely understood as it should be, especially among young economists.

One reason this matters, and may do so more in the future, is that:

"The Government's firm presumption is that after ... 2008 ... the system for assessing research quality ... will be mainly metrics-based." www.hm-treasury.gov.uk/media/20E/EA/bud06_ch3_192

Another reason is that many young economists (and some older ones) have become obsessed with not what their research says but rather with where it appears in print. This is understandable but potentially dangerous. It undermines the real reason we do our work: to understand the world and with luck help to improve it. When the criterion becomes 'did she get the article into publication A?' rather than 'did she do something interesting that anyone would ever care about?' then, in my opinion, values have become corroded and our purpose distorted.

As, roughly speaking, senior scholars get to make the rules for the young, they have a responsibility, I would say, to dissuade a but-did-she-get-it-in-the-AER? obsessive attitude. One should look for a middle ground between not caring at all about journals and caring about nothing else. Plainly, famous journals tend to be better than obscure journals. I mean by this that the average article in the former kind of journal is more lastingly interesting than in the latter. However, there is not as much information in this kind of statement as one might think.

Peters and Ceci (1982) did a nice experiment. It might not be allowed today. They resubmitted a collection of articles to the same journals that had recently published them. All the journals were distinguished academic research journals in psychology. The articles had originally been written by people from

leading universities. When resubmitted, Peters and Ceci took off those authors' names, and inserted fictitious ones, and gave them fictitious affiliations from unknown universities.

What happened? The great majority of these (already-published) articles were rejected by the referees. Only in a minority of cases, moreover, was it noticed by reviewers or editors that that paper had actually recently appeared in the same journal.

Another kind of evidence comes from Starbuck (2003). He is a distinguished researcher in management and business. For some years, Starbuck edited the famous Administrative Science Quarterly, which is the equivalent to the American Economic Review in economics or the Psychological Review in psychology. When he took over the editorship of ASQ, Starbuck did a study of the first 500 submissions he handled, and examined the referees' assessments of them (approximately 2 reports for each paper):

"The property of these reviews that struck me most vividly was their inconsistency. A surprisingly (to me) small fraction of the reviewers agreed with each other. Counting an Accept as 1, a Revise as a 0, and Reject as -1, I calculated the correlation. It was 0.12. Given the large sample, this correlation was statistically significant... but ... it was so low that knowing what one reviewer had said about a manuscript would tell me almost nothing about what a second reviewer had said or would say." Starbuck 2003, p. 346.

In our country, the 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. On such assessments will turn many taxpayers' pounds. Italy and Australia seem likely to follow the UK's example and introduce a form of state-run university assessment exercise. Partly because of the size of the undertaking, there will be pressure on members of these peer review panels to use journal labels (X is a 4* journal, Y a 2* journal, and so on) in a mechanical way to decide on the quality of articles. Rumours of this, and guesstimates of the key list of journals, are circulating. Similar forces are

visible elsewhere. Seglen (1997) notes the rising use of journal ratings as part of funding decisions in medical research. In the world of economics research, a fine Dutch research institute publishes a list of starred journals, ranked into categories of quality, to emphasise to its researchers that papers in certain journals should be viewed as of quality 'A' while others are of quality 'B'.

Some Data

Earlier this year I tried in a small way to explore the reliability of prestige labels in economics, to be reported in Oswald (2006). The project might be viewed as related to papers such as Laband and Tollison (2003), and newspaper articles such as Monastersky (2005). It seems complementary to work such as Laband (1990), Oswald (1991), Laband and Piette (1994), Johnson (1997), Kalaitzidakis et al (1999), Frey (2003), Seglen (1997), Coupe (2003), and Starbuck (2003, 2005), and is a minor contribution to the field of scientometrics (van Dalen and Henkens 2005, Sussmuth et al 2006). There is some link to the useful work of information-science researchers such as Oppenheim (1995), who have shown that, in the U.K., the departmental rankings that come out of the Research Assessment Exercise are correlated with ones that would have emerged from a citations-based departmental ranking.

Say we assume that after some decades the quality of a journal article is approximately known. One simple measure is that of impact as captured by the total ISI Web of Science citations the article has received (that is, the number of times the article has been quoted in later researchers' bibliographies).

Lots of research now uses citations to assess intellectual output and productivity. We know that U.S. professorial salaries are correlated with researchers' lifetime citations, and that citation counts are a fairly good predictor of Nobel and other prizes. And better universities are led by more highly-cited individuals. See, for example, Hamermesh et al (1982), Laband

(1990), Garfield and Welljams-Dorof (1992), Toutkoushian (1994), Moore et al (1998), Van Raan (1998), Thursby (2000), Bayers (2005) and Goodall (2006).

Of course, citations are a rough and highly noisy signal of quality. Survey articles tend to garner citations more easily than regular papers; there may be some pro-US bias in citations; sometimes papers may be cited because they are known to be wrong (though it is actually hard to find important examples of this); citation numbers are more open to manipulation than are publications figures; for some individuals, self-citations can cause problems; and so on.

However, citations are one measure of scholarly influence. They are more than a label, one might say. Unfortunately, if the impact-factors of journals become distorted over time, as is, I think, bound to happen as citations attract greater publicity and authors and editors try to manipulate citations totals, then the reliability of citations data will decline in the future.

I took a selection of economics journals from 1981 (there was no particular reason for this year, but it was a quarter of a century earlier, and I assumed I needed 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. Data on lifetime citations were then collected on each article. The raw data are summarized in Oswald (2006).

The mean lifetime cites across these six journals followed the broad pattern that might be expected. The prestige labels are, in a sense, correct. Mean numbers of cites per article published in the issue were: AER 68 cites; Econometrica 63 cites; JPubEcon 22; EJ 30; JIE 9; OBES 7. The top journals thus dominated. Similarly, median lifetime cites were: AER 23 cites; Econometrica 22 cites; JPubEcon 9; EJ 11; JIE 3; OBES 2.

But the variation of true quality -- as measured by cites -- was enormous. Consider, as a benchmark, the median number of cites to an article after a

quarter of a century. In the two elite journals here, it was approximately 22 cites. 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 principle, have appeared in the top journals. The answer was approximately 16% of the articles in lesser journals.

To put things in a starker way, I find in my data that it was far better to publish the top article in an issue of the Oxford Bulletin of Economics and Statistics (and probably lots of similar journals) than to publish the combined worst-4 articles in an issue of the American Economic Review. Yet few people on funding or promotion committees think about this possibility. My data are consistent with the theoretical argument of Starbuck (2005), who points out, using simple statistical parameterizations, that an error-ridden system would generate empirical patterns of the sort I document.

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 my raw numbers. 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 looks like prior insight. If we regress total cites, y , on publication-order in the journal, x , (that is whether the paper was first, second, third...eighteenth), we find a systematic relationship. Articles higher up a journal go on to acquire more cites over the ensuing quarter of a century. Hudson (2006), which was not available at the time the first draft of my work was done, finds equivalent results on the statistically significant role of the order of journal papers within an econometric equation explaining citations.

Thus if editors know something -- presumably using deep instinct -- then so may review panels in, say, a Research Assessment Exercise. Hence those

expert individuals should not feel obliged to apply journal labels in a mechanical way. They should use their judgment.

Conclusion

According to the data I collected, it is better, if the criterion is taken to be later citations, to publish the best article in an issue of the Oxford Bulletin of Economics and Statistics than to be the author of all four of the worst-4 articles in an issue of the American Economic Review. This does not mean that young scholars ought to ignore top journals, nor that research funders should (nor even that I will try to get my papers into little-known journals: I will not). But more maturity of outlook in the profession would surely be sensible.

The publication system in economics is full of error. It routinely pushes high-quality papers into medium-quality journals, and vice versa. Unless hiring committees, promotion boards, and funding bodies are aware of this fact, they are likely to make bad choices about whom to promote and how to allocate resources. It is our ideas, not our labels, that will be remembered after we are dead.

References

- Bayers, N.K. (2005). Using ISI data in the analysis of German national and institutional Research Output. *Scientometrics*, 62, 155-163.
- Coupe, T. (2003). The price is right: An analysis of best-paper prizes. Unpublished paper. National University of Kyiv-Mohyla, Ukraine.
- Frey, B.S. (2003). Publishing as prostitution? Choosing between one's own ideas and academic success. *Public Choice*, 116, 205-223.
- Garfield, E. and Welljams-Dorof, A. (1992). Of Nobel class: A citation perspective on high impact research authors. *Theoretical Medicine*, 13, 117-135.
- Goodall, A.H. (2006). Should research universities be led by top researchers and are they? A citations analysis. *Journal of Documentation*, 62, 388-411.
- Hamermesh, D.S., Johnson, G.E. and Weisbrod, A. (1982). Scholarship, citations and salaries: Economic rewards in economics. *Southern Economic Journal*, 49, 472-481.
- Hudson, J. (2006). Be known by the company you keep: Citations – quality or chance? Unpublished paper, University of Bath.
- Johnson, D. (1997). Getting noticed in economics: The determinants of academic citations. *The American Economist*, 41, 43-52.
- Kalaitzidakis, P., Mamuneas, T.P. and Stengos, T. (1999). European economics: An analysis based on publications in the core journals. *European Economic Review*, 43, 1150-1168.
- Laband, D.N. (1990). Is there value-added from the review process in economics? Preliminary evidence from authors. *Quarterly Journal of Economics*, 105, 341-352.
- Laband, D.N. and Piette, M.J. (1994). The relative impacts of economics journals – 1970-1990. *Journal of Economic Literature*, 32, 640-666.
- Laband, D.N. and Tollison, R.D. (2003). Dry holes in economic research. *Kyklos*, 56, 161-173.
- Monastersky, R. (2005). The number that is devouring science. *The Chronicle of Higher Education*, Issue: October 14.
- Moore, W.J., Newman, R.J. and Turnbull, G.K. (1998). Do academic salaries decline with seniority? *Journal of Labor Economics*, 16, 352-366.

- Oppenheim, C. (1995). The correlation between citation counts and the 1992 Research Assessment Exercise Ratings for British library and information science university departments. *Journal of Documentation*, 51, 18-27.
- Oswald, A.J. (1991). Progress and microeconomic data. *Economic Journal*, 101, 75-81.
- Oswald, A.J. (2006). An examination of the reliability of prestigious scholarly journals: Evidence and implications for decision-makers. To appear in *Economica*.
- Peters, J.P. and Ceci, S.J. (1982). Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behavioral and Brain Sciences*, 5, 187-255.
- Seglen, P.O. (1997). Why the impact factor of journals should not be used for evaluating research. *British Medical Journal*, 314, 497.
- Starbuck, W.H. (2003). Turning lemons into lemonade – Where is the value in peer reviews? *Journal of Management Inquiry*, 12, 344-351.
- Starbuck, W.H. (2005). How much better are the most-prestigious journals? The statistics of academic publication. *Organization Science*, 16, 180-200.
- Sussmuth, B., Steininger, M. and Ghio, S. (2006). Towards a European economics of economics: Monitoring a decade of top research and providing some explanation. *Scientometrics*, 66, 579-612.
- 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 Dalen, H.P. and Henkens, K. (2005). Signals in science – On the importance of signaling in gaining attention in science. *Scientometrics*, 64, 209-233.
- 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.