UNIVERSITY OF WAIKATO
Hamilton
New Zealand

The New Zealand Performance Based Research Fund and Its Impact on Publication Activity in Economics

David L. Anderson and John Tressler

Department of Economics

Working Paper in Economics 3/13

Revised August 2013

Corresponding Author
John Tressler
Economics Department
University of Waikato
Private Bag 3105
Hamilton
NEW ZEALAND, 3240

Tel: +64 (0) 7-838-4045
Email: tressler@waikato.ac.nz
Homepage:
http://wms-soros.mngt.waikato.ac.nz/personal/tressler

David L. Anderson
School of Business
Queen's University
Kingston, Ontario K7L 3N6
CANADA
Tel: 1-613-533-2362
Email: dla@queensu.ca
and
Waikato Management School
University of Waikato
Private Bag 3105
Hamilton, NEW ZEALAND
Email: anderson@waikato.ac.nz

Abstract

New Zealand's academic research assessment scheme, the Performance Based Research Fund (PBRF), was launched in 2002 with the stated objective of increasing research quality in the nation's universities. Evaluation rounds were conducted in 2003, 2006 and 2012. In this paper, we employ 22 different journal weighting schemes to generate output estimates of refereed journal paper and page production over three six year periods (1994-1999; 2000-2005 and 2006-2011). These time periods reflect a pre-PBRF environment, a mixed assessment period, and a pure PBRF research environment, respectively. Our findings indicate that, on average, research productivity, defined in either paper or page terms, has increased since the introduction of the PBRF. However, this outcome is due to a major increase in the quantity of papers and pages produced per capita that has more than off-set a decline in the quality of published outputs since the introduction of the PBRF. In other words, our findings suggest that the PBRF has failed to achieve its stated goal of increasing average research quality, but it has resulted in substantial gains in productivity achieved via large increases in the quantity of refereed journal articles.

Keywords

research measurement
PBRF
research quality
research assessment exercises

JEL Codes

A11, A14, C81, J24

Acknowledgements

We would like to thank John Gibson for providing a database of publications by New Zealand economists prior to 1999.

1. Introduction

Nation- wide assessments of university-based research are now relatively common. Arguably the best known and most influential of such reviews is the U.K.'s Research Assessment Exercise (RAE), first implemented in 1986.[1] Since then many other nations have, or are actively developing, their own research assessment schemes (OECD, 2010). In this paper we shall explore the impact of one such evaluation mechanism, New Zealand's Performance Based Research Fund (PBRF), on publication activity in the nation's university-based economics departments. The stated objective of the PBRF was to improve the quality of research in the nation's higher education sector,[2] and this was to be achieved, in large part, through financial incentives. More specifically, approximately 20 percent of the government's total annual operating grant to universities was to be distributed through the PBRF scheme, with individual universities receiving 60 percent of the available funding based on their relative per capita research performance and their PBRF eligible, full-time equivalent staffing levels (TEC, 2002).

The PBRF was one of the recommendations of the Tertiary Education Advisory Commission in 2001 and the details of the process to be used were not known until late 2002 (TEC, 2002). The scheme required all academics to submit to a peer assessment of their research over the most recent six year period. The funding implications of PBRF assessments for institutions were to be phased in over the period 2004- 2007. The first evaluation was conducted in 2003, based on research over the period 1997-2002. This was followed by a partial round in 2006, so called because all academics were given the option of opting out and using their 2003 grade as their 2006 result or being evaluated on their research output over the six year period starting on 1 January 2000. The third PBRF round was undertaken in 2012, with all research output over the period 2006-2011 being subject, once again, to peer review. At the time of writing, the 2012 results have not been released.

Before proceeding, we shall provide a brief outline of the PBRF evaluation scheme.[3] The PBRF is a mandatory, individually- based review of the research activities of all academics normally expected to undertake research. Researchers are assigned to one of 12 subject panels, and ultimately to one of 42 discipline categories, with grade assignments being made on the basis of peer assessment. Individual academics submit evidence portfolios (EP) that describe Research Output (RO), Peer Esteem (PE) and Contributions to

[1] The current version is known as the Research Excellence Framework (REF).

[2] References to desired improvements in research quality can be found in virtually all government documents outlining the scheme. For example: 'Cabinet has agreed that a Performance- Based Research Fund should be established to: increase the average quality of research', (TEC, 2002, p.7); '..primarily about quality, not quantity', TEC, 2011, p.104; and '.. encourage and reward research excellence', TEC, 2003, p. vii..

[3] For a more extensive discussion of the current version of the scheme, see TEC, 2011. Reviews and comments on the 2003 and 2006 versions of the PBRF can be found in Adams, 2008; Mathews and Sangster, 2009; Goldfinch, 2003; Boston, Mischewski and Smyth (2005); and in Bakker, Boston, Campbell and Smyth (2006).

the Research Environment (CRE). In describing Research Output (RO), researchers can include up to 34 research outputs, and identify 4 which they will describe more fully. It is important to note that research is broadly defined and covers areas such as refereed papers, conference papers, books and book chapters that have been produced over the six year evaluation period. However, it is widely recognized that for most disciplines, including economics, refereed papers are the primary yard-stick used in the evaluation process. Evidence on Peer Esteem (PE) is based on accolades received during the evaluation period, such as research grants, awards and honours, citations generated by previously published work, journal editorships and the like. Contributions to the Research Environment (CRE) include factors such as the researcher's role in doctoral supervision, development of research groups and programmes, and administrative roles in the area of graduate education. A researcher can refuse to submit an EP, but is then automatically given a score of zero. The EP is reviewed by the relevant expert panel and integer grades are assigned using a 0-7 scale for each of RO, PE and CRE. These numerical scores are then combined to give a Quality Evaluation (QE) score. In determining the Quality Evaluation score, the RO score has a weight of 70% and the PE and CRE scores a weight of 15% each.

Individual assessments are used in two ways. First, researchers are assigned to one of four categories: A, B, C and R (not active in research). Although the letter grade received by each researcher is supposed to be confidential, it is widely perceived that this is not the case, especially since high achievers frequently make their success known for pecuniary and ego reasons. At the discipline and institutional level, the letter grades are publicly available, and are frequently used in promotional material and in other forms of advertising by the 'winners'. Second, the letter grades are converted to a numerical score: R (0), C (1), B (3) and A (5). These scores are then doubled and are used to generate per capita scores for each discipline category for each university, and then aggregated across all categories to generate a per capita score for each university. The results are used in two distinct ways: first, as for letter grades, they are aggressively used by the leaders in each category in their marketing activities; and second, they are used, along with each university's PBRF eligible, full-time equivalent staff numbers, to allocate 60 percent of the available PBRF funding.[4]

These details of the PBRF are essential in understanding the intended effects of the scheme. For each individual academic the relevant peer-review panel has available to it information that can be used in assessing both the quality and quantity of the research undertaken. Since the peer evaluation process is essentially a black-box it is difficult to judge the weight given to quality versus quantity or how quality is assessed in determining the RO score.

Research Output (RO) and Quality Evaluation (QE) assessments have been treated as research quality measures in much of the discussion and analysis of PBRF. For example Smart (2008a) uses regression analysis to consider the determinants of research quality using the PBRF assessments for individual academics as the independent variable. Cinlar and

[4] The remaining 40 percent of PBRF funding is allocated as follows: 25 percent based on research degree completions, and 15 percent on external research funds attained.

Dowse (2008) use per capita scores in analysing the impact of PBRF in particular subject areas. Smart (2008b) considers 'Quality vs Impact' by comparing the PBRF assessments of institutions in particular subject areas with citation results. The impact of PBRF on research productivity across all sectors as a whole is studied in Smart (2009a) using ISI Web of Science (henceforth, ISI) refereed papers as a measure of research output. This research supports the view that research productivity increased substantially following the introduction of PBRF.

This literature is not clear on the distinction between 'quality' and 'output'. Since research outputs are not homogeneous, information on the quality of research produced is normally combined with assessments of quality to produce measures of output. Research productivity is then output per capita. Assessments of quality are involved in the implicit evaluations of peer review panels, or are based on quality indicators such as journal weighting schemes and citation analysis. In this sense the PBRF assessment of research is a measure of output and is referred to as a 'Research Output (RO)' score. [5], Determining 'quality' would then normally involve an assessment of the quality per unit of research or perhaps a comparison of quality assessed output with a measure of the quantity of research.

Although the objective of the PBRF is to increase the quality of research, relatively little evidence is available on the impact of PBRF on quality compared with quantity. The only paper that we are aware of that directly addresses this issue is Hodder and Hodder (2010). They studied the refereed paper output of three New Zealand business schools over the 2004-2008 period. The journal ranking scheme developed by the Australian Business Dean's Council (ABDC) was employed to assess quality changes. They summarize their findings as follows: 'While there are fluctuations in the proportion of articles in various grades of journals over the 2004- 2008 period, the overall tendency is towards more articles in journals of lower grades, i.e., the quantity of articles appears to be over-riding considerations of the quality of the journal' (p. 900). Indirect evidence on the quality issue is also provided by two other papers. Gibson, Tressler and Anderson (2008) use an academic labour market econometric model to show that the returns to quality declined and those to quantity increased in academic labour markets for economists following the introduction of PBRF up to the end of 2006. However the longer term results are open to question given the limited timeframe considered.[6] In Anderson, Smart and Tressler (2013) PBRF research output scores for individuals in Economics for the 2006 round are compared with research output measures using a variety of internationally recognised journal weighting schemes. This work indicates that PBRF assessments reflect only a moderate weight on quality and are not consistent with 'aggressive or high-powered' quality weighting schemes.

---

[5] PBRF terminology is not consistent here. When RO, CRE and PE scores are combined the result is referred to as a 'Quality Evaluation' (QE) score, even though this is an assessment of overall research output or contribution.

[6] In a related paper Smart (2009b) uses PBRF Quality Evaluation (QE) scores to consider the relationship between research quality and the probability of promotion for all those participating in the 2003 and 2006 rounds.

In this paper a detailed dataset covering the research output of academic economists between 1994 and 2012 is used to directly address the quality versus quantity question. To measure research we use a wide variety of alternative measures of output commonly used in the international literature on research evaluation. This enables us to consider both the quantity of measured research and its assessed quality per publication. In contrast to the above noted studies, we have access to publications over two PBRF rounds (2006 and 2012), and we employ a much broader range of alternative output measures allowing us to address the quality/quantity/productivity question in greater depth. At this point we must stress that we have restricted our analysis to one component of research; that is, we have ignored conference papers, books, chapters in books and working papers. However, the refereed article is widely employed in the economics literature as the primary measure of research success,[7] and it is widely perceived that the PBRF economics evaluation panel operates in a manner consistent with this view. Fortunately, we have access to both quantity and quality data with respect to refereed articles; therefore, we are able to generate estimates of productivity changes over the assessment period, and we can further disaggregate the results to reveal the role played by each variable. For quantity estimates, we have defined economics-relevant articles to be those contained in the approximately 1300 journals listed in *EconLit* as at 17 July 2012.

Selecting a proxy for article quality is more difficult. The economics literature abounds with journal ranking schemes, generally based on either adjusted or unadjusted citation counts. The rationale for using such journal weighting schemes (JWS) as a surrogate for article quality is the widespread view that citations are the best available measure for assessing an article's impact. Although problematic, we make the normal leap of faith, and equate impact with quality.[8] However, the JWS selection process is complicated by the fact that there are numerous journal weighting/ranking schemes, and most differ in several important ways. For example, decisions must be made with respect to: treatment of citations from multi-authored papers; treatment of self- citations (accept or reject); selection of an appropriate citation database; determination of an appropriate time period for the collection of cites; the selection of the type and range of journals eligible to send and receive citations; and acceptance or rejection of the premise that all cites are of equal value; and, if they are not judged to be equal, how should cites be adjusted. One can also utilize journal based weighting schemes that are based on perceptions of experts. As above, perception schemes vary in many ways, but most importantly in the biases held by voters.

Before proceeding, we must address an obvious omission: our failure to include direct citation counts and derivatives therefrom such as the g and h index. The reason is simple: the lag structure in economics (and, more generally, the social sciences) between acceptance of an article and the generation of a meaningful stream of citations is such that the results are unreliable in the context of an individual-based, six year cycle, research assessment scheme

[7] In fact, virtually all articles in the rankings literature in economics focus on refereed papers to the exclusion of all other forms for research, largely for pragmatic reasons (data availability) and for a belief that such output is of primary importance in tenure, promotion and hiring decisions.

[8] For a critique of this assumption, see Beed and Beed (1996).

(Tressler and Anderson, 2012).[9] Therefore, for purposes of this study, the direct citation approach is rejected in favour of the indirect approach.

2. Data Sources

In order to explore the quantity/quality issue in a rigorous fashion, we assembled three datasets covering all staff employed by New Zealand's eight universities as at 2000, 2006 and 2012. For each reference period we collected publications generated over the preceding six years in order to replicate the assessment periods of the PBRF rounds of 2006 and 2012 (2000-2005 and 2006-2011, respectively). The dataset based on the year 2000 staff lists, covering publications over the period 1994-1999, provides us with information on the incentive structure facing New Zealand's academic economists prior to public discussion of a PBRF-like research evaluation scheme.[10] The first two staff lists were derived from university calendars, and the 2012 lists were obtained from departmental websites on 15 April 2012. For each staff member, publication lists were compiled primarily from the EconLit database, supplemented by reference to the Research Papers in Economics RePEc) database and individual CVs. We followed the prevailing practice of allocating shares in multiple-authored papers on the basis of the 1/n rule, and adopted the same rule in allocating pages to individual authors. We again followed convention, and adjusted page counts to reflect size variation by converting all pages into *American Economic Review (AER)* equivalents.[11] Hence, we have two measures of output that we shall use throughout this paper: share adjusted papers and share and size adjusted pages.[12]

We do not intend to suggest which journal weighting scheme (JWS) is best, but instead will use a large number of currently available measures to see if output changes are primarily due to quantity or quality factors, and also to see if the results differ across competing measures. Since the JWS selection process is arbitrary, we employed twenty-two schemes

[9] In another study, based on ISI data, Anderson and Tressler (2013) found that over a six year period 52 percent of New Zealand economists received a 'g' and 'h' score of zero, and 92 and 89 percent received 'h' and 'g' scores of two or less, respectively. It is our view that these results fail to differentiate in a meaningful way between low and high producers.

[10] For purposes of this study, we have ignored the 2003 PBRF round. Although the official scheme was not announced until late 2002, it was common knowledge from 1999 onwards that the government intended to move towards some form of research evaluation. Hence, it is possible that some academics adjusted their behaviour accordingly. However, the general view is that most academics were taken by surprise when the PBRF was announced (for example, see Hazledine and Kurniawan, 2005, p.173). Nevertheless, we have chosen to adopt a conservative approach and create a dataset that precedes any possible behavioural change.

[11] We wish to thank Joseph Macri and the late Dependra Sinha for providing us page conversion factors for approximately 500 journals. For all other journals, we used 0.72 as ratio of page size in 'other' journals to an average *AER* page. Gibson (2000) provides justification for this estimate.

[12] For a review of the conventional literature with respect to research measurement in economics, see Macri and Sinha (2006).

in an attempt to capture a wide range of factors (see Appendix 1 for a complete list and a brief outline of associated characteristics). For example, we selected schemes based on the ISI, Scopus and RePEc databases. We also selected schemes (using one of the three aforementioned databases) that are primarily based on simple impact factors (that rely heavily on the work of Garfield, 1972) and others based in large part on the recursive adjustment approach initially used in the journal ranking literature by Liebowitz and Palmer (1984). It should also be noted that we also utilized four perception-based schemes and a base-line measure that grants equal weight to all journals and all articles and pages therein.

3. Results

Let us now explore unadjusted research output changes over our three time periods (BASE (1994-1999), PBRF1 (2000-2005) and PBRF2 (2006-2011). From Table 1 it can be seen that Unadjusted Papers/Capita increased by 78 percent from BASE to PBRF2. The increase in share adjusted papers is somewhat lower at 32 percent indicating an increase in the number of authors per paper. This is undoubtedly due, in part, to discipline wide trends, but may also be attributable to research management strategies to shift more staff from the non-active to the research active group (Hodder and Hodder, 2010). Finally, from Table 1, note that un-weighted, but share and size adjusted page output per capita increased by over 63 percent from the BASE period to PBRF2. In summary, it is clear that from a productivity perspective, albeit with all paper and page output deemed to be of equal value, the PBRF programme appears to have been a success. However, not everything else is equal in the publication realm, and to address this issue we shall now employ the previously referenced 22 journal weighting schemes (JWS).

Table 1: Ratio of Output Between Time Periods
Various Categories

	Number of Papers	Papers Per Capita	Share Adjusted Papers Per Capita	Size and Share Adjusted Pages Per Capita
06-11: 94-99	2.08	1.78	1.32	1.63
06-11: 00-05	1.40	1.41	1.17	1.23
00-05: 94-99	1.48	1.26	1.13	1.32

Table 2. Ratio of Standardized Scores Between Various Time Periods
Simple Average, 22 Journal Weighting Schemes

Time Periods	Adjusted Papers Per Capita	Adjusted Pages Per Capita	Score Per Adjusted Paper	Score Per Adjusted Page
06-11: 94-99	109.27	131.72	82.68	81.02
06-11: 00-05	95.08	113.83	83.91	85.79
00-05: 94-99	114.54	114.98	98.20	93.84

In constructing Table 2, the BASE period scores for all 22 JWS have been set at 100.0, and the output for other time periods adjusted accordingly. The results presented in Table 2 are the average scores across all 22 schemes. First, note that on average, the number of share adjusted weighted papers (henceforth **Adjusted Papers**) per capita increased by 9 percent in PBRF2 relative to the BASE period. This is in comparison to a 32 percent increase in un-weighted, but share adjusted papers per capita. The corresponding numbers for share and size adjusted weighted pages (henceforth **Adjusted Pages**) and un-weighted, share and size adjusted pages are 32 and 63 percent, respectively. This suggests that much of the increase in output per capita between BASE and PBRF2 is due to a shift towards lower quality journals. Nevertheless, productivity, as measured in either **Adjusted Papers** or **Adjusted Pages** terms, has increased over the life of the PBRF, allowing one to argue that the PBRF scheme has been a success from an economic efficiency perspective. However, the official objective of the PBRF is not to improve productivity, but to increase the quality of research. Let us now turn our attention to this issue.

To measure quality, we now shift our yard stick from output per capita to score per paper and per page. Once again, refer to Table 2. Note that the average score per **Adjusted Paper** in period PBRF2 is only 83 percent of that prevailing in the BASE period. Interestingly, the rate of decline in average score per **Adjusted Page** is even greater: from 100 in BASE, to 81 in PBRF2. In other words, the score associated with the average **Adjusted Page** declined by 19 percent between BASE and PBRF2. Therefore, our aggregate results suggest that the PBRF scheme has not raised quality, but has encouraged the publication of research in lesser ranked journals. This may have resulted from the increased productivity of staff who typically publish in lower ranked journals, even though individual academics are encouraged to publish in better journals. Alternatively, for some academics increasing the quantity of research output may have been an easier road to success than the pursuit of quality.

Table 3. Increases/Decreases in Scores Between Various Time Periods

A. Change in Score from 94-99 to 06-11 for 22 Journal Weighting Schemes

	Adjusted Paper Per Capita	Adjusted Page Per Capita	Score Per Adjusted Paper	Score Per Adjusted Page
Decline	6	2	20	20
No Change	0	0	1	1
Rise	16	20	1	1

B. Change in Score from 00-05 to 06-11 for 22 Journal Weighting Schemes

	Adjusted Paper Per Capita	Adjusted Page Per Capita	Score Per Adjusted Paper	Score Per Adjusted Page
Decline	1	2	12	19
No Change	0	0	1	1
Rise	21	20	9	2

The above analysis has been based on the average scores over 22 JWS. Therefore, it is reasonable to ask if our results are dependent on the type of weighting scheme employed in the analysis. As shown in Table 3,[13] this does not appear to be the case. From a productivity perspective, most, but not all JWS moved in the same direction for both **Adjusted Paper** and **Adjusted Page** measures. More specifically, 16 of 22 schemes are associated with an increase in **Adjusted Paper** productivity between the 1994-1999 and 2006- 2011 periods. The corresponding figure for **Adjusted Page** productivity is 20 of 22. Note that the change from PBRF1 to PBRF2 suggests that productivity has continued to increase over time: the corresponding estimates are 21 of 22 and 20 of 22 for **Adjusted Paper** and **Adjusted Page** estimates, respectively. In summary, most, but not all JWS lead to the conclusion that productivity has increased during PBRF1 and, especially PBRF2, relative to the preceding six year period (BASE).

Let us now look at what the disaggregated results imply about the quality issue. Once again, refer to Table 3. Before proceeding, note that our EQUAL scheme will, by definition, indicate no change in quality over time. The results are strongly supportive of the view that the PBRF scheme has failed to achieve its stated objective. In fact, 20 of 21 schemes (excluding EQUAL) exhibited a decline in **Adjusted Paper** and **Adjusted Page** quality over the BASE-PBRF2 period. Hence, the JWS selection process does not appear to be of great importance: virtually all journal weighting schemes lead to the same result: productivity has improved and quality has declined for both **Adjusted Paper** and **Adjusted Page** indicators under the PBRF regime.

However, reference to Appendix 2B does shed some light on the importance of the JWS selection process with respect to the productivity and quality issue. Although admittedly arbitrary, we have grouped twenty of our twenty-two journal weighting schemes into three broad categories: those based in large part on recursive citation adjustment processes wherein the value of a citation depends upon the citing journal; those based in large part on simple impact factors, albeit, sometimes with adjustments to reflect sub-discipline citing habits, etc.., and schemes based on the perceptions of 'experts'. For purposes of discussion, we denote the groups as Recursive, Impact Factor and Perception Schemes. Based on Recursive Schemes, productivity drops slightly when based on **Adjusted Papers** but rises by 15 percent for **Adjusted Pages** over the BASE to PBRF2 period. However, the quality scores for both **Adjusted Papers** and **Adjusted Pages** decline by 27 and 29 percent, respectively. On the other hand, the Impact Factor schemes yield quite different results: the PBRF scheme would be judged to have led to a major increase in productivity, with the average increase being 23 percent and 52 percent for **Adjusted Papers** and **Adjusted Pages**, respectively. On average, quality scores based on Impact Factor Schemes also decrease over time, but much more modestly than under the Recursive Schemes.

Our third category, Perception Schemes, yields, on average, results that are roughly mid-point between those associated with the Recursive and Impact Factor Schemes: productivity increases by roughly 15 percent and 38 percent for **Adjusted Papers** and **Adjusted Pages**,

[13] See Appendix 2A for supporting information.

respectively between 1994-1999 (BASE) and 2006-2011 (PBRF2). Corresponding figures for quality are -13 and -15 percent. It should be noted that three members of the Perception Set (ESA, ERA and ABCD10) are Australian based. Given the close ties between Australia and New Zealand, these schemes are the best proxies available for reflecting the importance of regional and national issues**.**

4. Policy Implications and Conclusions

Our findings suggest that the PBRF's stated objective of increasing research quality has not been achieved. On the other hand, productivity, or output per capita, has improved from both an **Adjusted Paper** and **Adjusted Page** perspective, as the increase in quantity has more than compensated for the decline in quality. It is important to note that we found these results to be relatively stable across the 22 journal weighting schemes employed in our analysis.

At this point we digress somewhat and address the following questions: why and how did productivity increase and quality decline during the PBRF regime. The 'why' is subject to speculation, but it appears as if research managers and individual researchers were able to obtain enough information about the black-box used by the discipline assessment panel to realize that, at the very least, if everything else was equal, quantity mattered. Hence, previously inactive staff were encouraged to produce some *EconLit* recognized output, since even a small amount of output was deemed to be sufficient to move such individuals from an R to a C, resulting in an increase in the unit's per capita score. In the 2006 evaluation 65.34% of eligible staff in Economics were rated in the C, C(NE), R or R(NE) categories. [14] In contrast only 6.9% of staff received an A while 38.35 were rated B. Many of those who received a low rank in 2006 would have had few or no journal publications. For these academics the best strategy for achieving an increased rank in 2012 was to increase the number of publications in relatively low-ranked journals rather than attempting a major increase in the quality of research, or publications in high quality journals. Even for staff at higher ranks, maintaining or increasing research productivity at the prior quality level may have been a less risky strategy than aiming for a significant increase in the quality of research. It is also possible that research managers recognized that some previously research- inactive staff were not likely to change their ways, and that some of these people should be encouraged to depart or be reclassified to non- assessed positions (read: teaching fellows and the like). Once again, this action leads to an improvement in per capita scores.

Let us now look at the actual changes in staffing patterns over the 2000, 2006 and 2012 periods to see if they align with the above mentioned score enhancing strategies. From Table 4, it is clear that the rank composition of the staff changed rather substantially over time, especially at the high (Professor) end. In fact, the percentage of Professors doubled from

[14] For the 2006 PBRF round, a new category for each of the C and R grades was added to reflect the fact that it was difficult for new and emerging scholars to compete successfully under the PBRF guidelines. The new categories were labeled as C(NE) and R(NE), with NE standing for 'new and emerging'.

2000 to 2012 (11 to 22 percent), and the percentage of Associate Professors also increased, albeit, at a more modest rate (from 16 to 19 percent). At the other end of the spectrum, the proportion of Senior Lecturers and Lecturers declined from 50 and 23 percent to 40 and 19 percent, respectively. Given that higher ranked academics are expected to be more productive than lower ranked staff, this change should result in increased output per capita (everything else constant). Although it is not clear why the resulting increase in per capita research output favoured quantity over quality, it is reasonable to suggest that hiring and promotion committees perceived the PBRF reward system to be one that rewarded high quantity/good quality producers over low quantity/high quality producers. We also suggested that one of the easiest ways for research administrators to increase absolute scores was to move non-producers to the active researcher category (expected to change letter grades from R (not research active) to C (acceptable). According to our databases, the proportion of academic staff that can be defined as research active (those individuals publishing at least one paper, in whole or in part, over a six year period in an *EconLit* listed journal) increased from 71 percent in BASE (1994- 1999) to 79 percent in PBRF1 (2000- 2005) and 83 percent in PBRF2 (2006-2011).

Table 4. Economists, New Zealand Universities, Percent by Rank

	Year		
Rank	2000	2006	2012
Professor	11.1	18.2	21.9
Associate Professor	16.2	14.6	19.0
Senior Lecturer	49.6	45.3	40.1
Lecturer	23.1	21.9	19.0

In our opinion, the above noted outcomes are consistent with the design characteristics of the PBRF. The scheme is based on a productivity framework: the unit of account is research output per capita. Therefore, even if the discipline assessment groups favour quality (however defined) over quantity, it is reasonable to expect that if everything else is equal, increased quantity will result in a higher score. This seems to be the prevailing view of how the peer review 'black box' operates. It has led research managers to change the composition of their work force to employ more PBRF-friendly researchers, and has also led to efforts to improve the productivity of previously low or non-producers on the premise that some output is better than no output, and that some research output invariably leads to a movement away from the dreaded R classification.

In conclusion, our findings suggest that the PBRF scheme has failed to achieve its intended effect (increased quality); however, it has generated a socially worthwhile

unintended effect (increased productivity).[15] It should be noted that our results are derived from a comprehensive database – 100 percent coverage of academic staff at the three census dates and, for each researcher, an assessment of all publications in *EconLit* listed journals. However, we must qualify our findings in three ways. First, our work has undoubtedly been aided by the fact that New Zealand is a small country with only eight university-based economics departments. Second, our work has addressed only one of forty-two discipline groups covered by PBRF, and it is possible that economics may not be representative of all or even most disciplines. Finally, our results are obviously scheme-specific, and should not be interpreted to imply that research assessment schemes, in general, are biased against quality improvements.

References

Abelson, P. (2009). The Ranking of Economics Journals by the Economic Society of Australia. *Economic Papers*. 28(2): 176- 180.

Adams, J. (2008). Strategic Review of the Performance-Based Research Fund: The Assessment Process. Evidence Ltd, Leeds, UK.

Anderson, D.L, Smart, W., and Tressler, J. (2013). Evaluating Research- Peer Review Team Assessment and Journal based Bibliographic Measures: New Zealand PBRF Research Output Scores in 2006. *New Zealand Economic Papers*, forthcoming.

Anderson, D.L. and Tressler, J. (2013). The Relevance of the 'h' and 'g' Index to Economics in the Context of a Nation- wide Research Evaluation Scheme: The New Zealand Case. *Economic Papers*, forthcoming.

Bakker, L, Boston, J., Campbell, L., and Smyth, R. (2006), eds. Evaluating the Performance-based Research Fund: Framing the Debate. Institute of Policy Studies, Wellington, New Zealand.

Bauwens, L. (1998). A New Method to Rank University Research and Researchers in Economics in Belgium. Unpublished paper, CORE, Universite Catholique de Louvain, Louvain, Belgium.

Beed, C. and Beed, C. (1996). Measuring the quality of academic journals: the case of economics. *Journal of Post Keynesian Economics.* 18(3): 369-396.

Boston, J., Mischewski, B. and Smyth, R. (2005). Performance-Based Research Fund- Implications For Research In The Social Sciences and Social Policy. *Social Policy Journal of New Zealand.* Issue 24: 55-84.

Cinlar, N. and J. Dowse (2008). Staffing and Performance Trends in Research Subject Areas. New Zealand Tertiary Education Commission, Wellington, New Zealand.

Combes, P.P. and Linnemer, L. (2010). Inferring Missing Citations: A Quantitative Multi-Criteria Ranking of all Journals in Economics. Groupement de Rechereche en Economie Quantitative d'Aix Marseille (GREQAMJ), document de Travail, no 2010-28.

[15] However, concerns have been expressed about possible negative impacts of the PBRF on teaching, collegiality and working conditions. For example, see Cupples and Pawson (2012): Waitere, Wright, Tremaine, Brown and Pause (2011); Middleton (2009); and Duncan (2007).

Coupe, T. (2003). Revealed Performances: Worldwide Rankings of Economists and Economics Departments, 1990-2000. *Journal of the European Economic Association.* 1(6): 1309-1345.

Cupples, J. and Pawson, E. (2012). Giving an account of oneself: The PBRF and the neoliberal university. *New Zealand Geographer*. 68: 14-23.

Duncan, G. (2007). What's wrong with the PBRF? *Politics of Happines.* http://grantduncan.blogspot.com, 16 August.

Garfield, E. (1972). Citation Analysis as a Tool in Journal Evaluation. *Science*, 178(4060): 471-479.

Gibson, J., Anderson, D.L. and Tressler, J. (2008). Do Research Assessment Exercises Raise the Returns to Publication Quality? Evidence from the New Zealand Market for Academic Economists. Working Paper 08/11, University of Waikato, Hamilton, New Zealand, www.RePEc.org.

Gibson, J. (2000). Research Productivity in New Zealand University Economics Departments; Comment and Update. *New Zealand Economic Papers.* 34(1): 73-88.

Goldfinch, S. (2003). Investing in Excellence? The Performance-based Research Fund And Its Implications For Political Science Departments in New Zealand. *Political Science*. 55(1): 39-53.

Hazledine, T. and Kurniawan, C. (2005). Efficiency of New Zealand Universities and the Impact of the Performance Based Research Fund. *New Zealand Economic Papers.* 39(2): 153-179.

Hodder, A.P.W. and Hodder, C. (2010). Research Culture and New Zealand's performance-based research fund: some insights from bibliographic compilations of research outputs. *Scientometrics.* 84: 887-901.

Kalaitzidakis, P., Mamuneas, T. and Stengos, T. (2011). An Updated Ranking of Academic Journals in Economics. *Canadian Journal of Economics.* 44(4): 1525-1538.

Kalaitzidakis, P., Mamuneas, T. and Stengos, T. (2010). An Updated Ranking of Academic Journals in Economics. Working Paper 9/2010, Economics Department, University of Guelph, Guelph, Canada.

Kalaitzidakis, P., Mamuneas, T. and Stengos, T. (2003). Ranking of Academic Journals and Institutions in Economics. *Journal of the European Economic Association.* 1(6): 1346-1366.

Kodrzycki, Y. and Yu, P. (2006). New Approaches to Ranking Economics Journals. *B.E. Journal of Economic Analysis and Policy: Contributions to Economic Analysis and Policy.* 5(1): Article 24.

Laband, D. and Piette, M. (1994). The Relative Impact of Economics Journals. *Journal of Economic Literature*. 3(2): 640-666.

Liebowitz, S.J. and Palmer, J.P. (1984). Assessing the Relative Impact of Economics Journals. *Journal of Economic Literature.* 22(1): 77-88.

Macri, J and Sinha, D. (2006). Rankings Methodology for International Comparisons of Institutions and Individuals: An Application to Economics in Australia and New Zealand. *Journal of Economic Survey.* 20(1): 111-156.

Mason, P., Steagall, J. and Fabritius, M. (1997). Economics Journal Rankings by Type of School: Perceptions versus Citations. *Quarterly Journal of Business and Economics.* 36(1): 69-79.

Mathews, M.R. and Sangster, A. (2009). Exporting the RAE: adoption of similar practices in Australia and New Zealand. *Asian Review of Accounting.* 17(2): 115-135.

Middleton, S. (2009). Becoming PBRF-ABLE: Research Assessment and Education in New Zealand. In: Besley, T., ed. Assessing the Quality of Educational Research in Higher Education: International Perspectives. Sense Publishers, Rotterdam, pp. 193-208.

OECD (2010). Performance-based Funding for Public Research in Tertiary Education Institutions. Workshop Proceedings. OECD Publishing, http://dx.doi.org/10.1787/9789264094611-en.

Smart, W. (2009a). The Impact of the Performance-Based Research Fund on the Research Productivity of New Zealand Universities. *Social Policy Journal of New Zealand.* Issue 34: 1-18.

Smart, W. (2009b). Promoting Quality Research. Tertiary Sector Performance Analysis and Reporting, Strategy and System Performance, New Zealand Ministry of Education, Wellington, New Zealand.

Smart W. (2008a). Research Par Excellence. Tertiary Sector Performance Analysis and Reporting, Strategy and System Performance, New Zealand Ministry of Education, Wellington, New Zealand.

Smart W, (2008b). Quality vs Impact. Tertiary Sector Performance Analysis and Reporting, Strategy and System Performance, New Zealand Ministry of Education, Wellington, New Zealand.

TEC (2011). Performance Based Research Fund: Quality Evaluation Guidelines. Tertiary Education Commission. Wellington, New Zealand.

TEC (2003). Performance-Based Research Fund: A guide for 2003. Tertiary Education Commission, Wellington, New Zealand.

TEC (2002). Investing in Excellence: The Report of the Performance-Based Research Fund Working Group. Tertiary Education Commission. Wellington, New Zealand.

Tressler, J and Anderson, D.L. (2012). Using Citations to Measure Research Output in New Zealand Economics Departments: The Problem of 'Long and Variable' Lags. *Agenda: A Journal of Policy Analysis and Reform.* 19(1): 17-37.

Waitere, H.J., Wright, J., Tremaine, M., Brown, S., and Pause, C.J. (2011). Choosing whether to resist or reinforce the new managerialism: the impact of performance-based research funding on academic identity. *Higher Education Research and Development.* 30(2): 205-217.

Appendix 1

Description of Journal Weighting Schemes

1) **LP84.** Source: Liebowitz and Palmer (1984. Database: ISI. Time Period: 1980 cites to articles published 1975-1979. Number of Journals: 108. Methodology: Recursive Adjustment.

2) **LP94**. Source: Laband and Piette (1994). Database: ISI. Time period: 1990 cites to articles published 1985-1989. Number of Journals: 130. Methodology: Recursive Adjustment.

3) **KMS2003.** Source: Kalaitzidakis, Mamuneas and Stengos (2003). Database: ISI. Time Period: 1998 cites to articles published 1994-1998. Number of Journals: 143. Methodology: Recursive Adjustment.

4) **KMS2010.** Source: Kalaitzidakis, Mamuneas and Stengos (2010). This working paper, with journal rankings, but minus absolute scores, was ultimately published, under the same title, in 2011 in the *Canadian Journal of Economics* 44(4): 1525-1538. Database: ISI. Time Period: average of citation counts to selected journals for preceding 10 years, starting in 2003 and continuing to 2008. Number of Journals: 209. Methodology: Recursive Adjustment.

5) **KYEI.** Source: Kodrzycki and Yu (2006). Database: ISI. Time Period: 2003 ISI citation data for papers published over the period 1996- 2003. Number of Journals: 181. Methodology: Recursive Adjustment with modification for sub-discipline reference intensity.

6) **KYOI.** Source: same as for KYEI above. Description same as for KYEI with one exception: citations collected from all Social Science journals in the ISI database to selected economics journals.

7) **KYPI.** Source: same as for KYEI above. Description same as for KYEI with one exception: citations counts restricted to those from an arbitrarily defined set of policy journals.

8) **CLd.** Source: Combes and Linnemer (2010). Database: ISI, Google Scholar and EconLit. Time Period: 2005, 2006 and 2007 for articles published in the preceding 5 years. Number of Journals: 1168. Methodology: Various- including simple impact factors, field of specialization normalized indexes and h-index. Under the CLd scheme, all 1168 journals are assigned to six groups with weights ranking from 100 to 3.125.

9) **CLm.** Source: same as CLd above. CLm differs from CLd in that journal values are adjusted to reflect a medium degree of convexity.

10) **CLh.** Source: same as CLd above. CLh differs from CLd in that journal values are adjusted to reflect a high degree of convexity.

11) **Gibson.** Source: Gibson (2000). Database: author constructed, primarily from *EconLit*. Time Period: i) 1996- 1998; and ii) life-time EconLit publications of selected researchers. Number of Journals: explicit weights for 71 journals; all other *EconLit* listed journals received an equal, non-zero equal to $1/20^{th}$ of top journals. Methodology: Econometric (Ordered Logit).

12) **CoupeIF.** Coupe (2003). Complete set of weights obtained from: http//homepages.ulb.ac.be/~tcoupe/update/journals.html (downloaded on 20 August 2007). Database: ISI. Time Period: 1994- 2000. Number of Journals: 273. Methodology: Simple Impact Factor (average of 2 year Impact Factors from 1994 to 2000).

13) **Bauwens.** Bauwens (1998). Database: EconLit and ISI. Time Period: 1996. Number of Journals: 617. Methodology: Simple Impact Factor- average of short-term (2 Year) and long-term (life-time of journal).

14) **Scopus SJR.** Source: Scopus/Economics, Econometrics and Finance Journal List, SJR-SCImago Journal Rank, www-scopus-com. Downloaded 17-19 October 2012. Database: Scopus. Number of Journals: +/- 950 (Economics and Finance). Methodology: Impact Factor, with adjustment for importance of citing journal.

15) **Scopus SNID.** Source: same as Scopus SJR above. SNID- Source Normalized Impact per Paper. Database: Scopus. Number of Journals: +/- 950 (Economics and Finance). Methodology: Impact Factor, with adjustment for reference intensity across and within disciplinary groups.

16) **RePEcsif.** Source: Research Papers in Economics (RePEc), IDEAS/RePEc Simple Impact Factors for Journals, www.ideas.repec.org/top/top.journals.simple.html, downloaded on 6 May 2012. Database: ISI. Time Period: life time of journal. Number of Journals: 984. Methodology: Simple Impact Factor.

17) **JCRIF11.** Source: ISI Web of Knowledge, Journal Citation Reports (2011), Impact Factor (2 Year). Downloaded on 8 October 2012. Database: ISI. Time Period: 2010 cites to articles published 2008-2009. Number of Journals: 321. Methodology: Simple Impact Factor.

18) **MSF.** Source: Mason, Steagall and Fabritius (1997). Database: Author conducted survey. Time Period: 1992- 1993. Number of Journals: 142. Methodology: Perceptions of respondents.

19) **ESA.** Source: Abelson (2009). For itemized rankings, see: http://www.ecosoc.org.au/cc/publications/. (Valid at time article published). Database: Survey conducted by the Economic Society of Australia. Time Period: 2007. Number of Journals: 890. Methodology: Perceptions of respondents.

20) **ERA.** Source: Excellence for Research in Australia, 2010 Final Journal Rankings, www.arc.gov.au/era. For a more user friendly presentation journal rankings, see John Lamp, Deakin University: http://lamp.infosys.deakin.edu.au/era. Downloaded on 9 March 2010. Database: Unknown. Time Period: 2010. Number of Journals: 640 (Economics). Methodology: Perceptions of Committee Members (supported by submissions also based on perceptions).

21) **ABDC10.** Source: Australian Business Deans Council Journal Quality List (2010). www.abdc.edu.au, downloaded on 12 October 2012. Database: Unknown. Time Period: 2010. Number of Journals: 907 (Economics). Methodology: Perceptions of Committee Members (supported by submissions, also based on perceptions).

22) **EQUAL.** All journals listed in *EconLit* (as at 17 July 2012), and articles and pages therein, are deemed to be of equal value. Number of Journals: +/- 1300.

Appendix 2A

Comparison of Standardized Scores, 06-11 relative to 94-99

(94-99 Scores Set to 100.0)

Journal Weighting Scheme	Adjusted Papers/ Capita	Adjusted Pages/ Capita	Score/ Adjusted Paper	Score/ Adjusted Page
LP84	82.36	104.71	62.31	64.44
LP94	68.75	73.81	52.00	45.33
KMS2003	77.25	82.97	58.48	50.93
KMS2010	86.54	104.27	65.38	64.10
KYEI	101.70	125.33	77.02	77.14
KYOI	102.72	126.46	77.82	77.80
KYPI	115.34	143.08	87.32	88.01
CLd	114.57	138.98	86.68	85.37
CLm	116.82	138.26	88.40	84.93
CLh	96.43	111.35	72.96	68.43
Gibson	110.27	122.34	81.42	76.28
CoupeIF	112.07	136.87	86.51	84.17
Bauwens	117.63	141.07	88.84	86.49
Scopus SJR	123.74	150.67	94.04	92.60
Scopus SNID	129.68	159.83	97.54	98.47
RePEcsif	122.40	156.08	92.47	95.90
JCRIF11	135.26	165.02	103.38	101.97
MSF	98.53	118.68	74.20	72.68
ESA	122.14	147.22	92.53	90.40
ERA	124.91	150.32	94.56	92.27
ABDC10	112.39	137.83	85.08	84.76
EQUAL	132.37	162.76	100.00	100.00
Average	*109.27*	*131.72*	*82.68*	*81.02*

Appendix 2B

Average Score by Methodological Grouping

RAS Category: Recursive Adjustment

Journal Weighting Scheme	Adjusted Papers Per Capita	Adjusted Pages Per Capita	Score Per Adjusted Paper	Score Per Adjusted Page
LP84	82.36	104.71	62.31	64.44
LP94	68.75	73.81	52.00	45.33
KMS2003	77.25	82.97	58.48	50.93
KMS2010	86.54	104.27	65.38	64.10
KYEI	101.70	125.33	77.02	77.14
KYOI	102.72	126.46	77.82	77.80
KYPI	115.34	143.08	87.32	88.01
CLd	114.57	138.98	86.68	85.37
CLm	116.82	138.26	88.40	84.93
CLh	96.43	111.35	72.96	68.43
Average	*96.25*	*114.92*	*72.84*	*70.65*

RAS Category: Impact Factor

Journal Weighting Scheme	Adjusted Papers Per Capita	Adjusted Pages Per Capita	Score Per Adjusted Paper	Score Per Adjusted Page
CoupeIF	112.07	136.87	86.51	84.17
Bauwens	117.63	141.07	88.84	86.49
Scopus SJR	123.74	150.67	94.04	92.60
Scopus SNID	129.68	159.83	97.54	98.47
RePEcsif	122.40	156.08	92.47	95.90
JCRIF11	135.26	165.02	103.38	101.97
Average	*123.46*	*151.59*	*93.80*	*93.27*

RAS Category: Perception

Journal Weighting Scheme	Adjusted Papers Per Capita	Adjusted Pages Per Capita	Score Per Adjusted Paper	Score Per Adjusted Page
MSF	98.53	118.68	74.20	72.68
ESA	122.14	147.22	92.53	90.40
ERA	124.91	150.32	94.56	92.27
ABDC10	112.39	137.83	85.08	84.76
Average	*114.49*	*138.51*	*86.59*	*85.03*