The Impact on Research Quality of Performance-Based Funding: The Case of New Zealand's PBRF Scheme

Robert A. Buckle and John Creedy

WORKING PAPER 03/2018
March 2018

Working Papers in Public Finance

VICTORIA
UNIVERSITY OF WELLINGTON
TE WHARE WĀNANGA
O TE ŪPOKO O TE IKA A MĀUI

Chair in Public Finance
Victoria Business School

The Impact on Research Quality of Performance-Based Funding: The Case of New Zealand's PBRF Scheme*

Robert A. Buckle and John Creedy

Victoria University of Wellington

NOTE: This paper corrects (October 2018) some numbers reported in the earlier version (in Tables 1, 5 and 6)

Abstract

This paper discusses the impact on research quality of New Zealand universities of the Performance-Based Research Fund from 2003 to 2012. This is a peer-review process involving assessment of individual researchers. The contribution to improvement in research quality arising from transitions among research quality categories and entrants and exits of individuals are identified. A substantial component of change has been the removal of non-research active staff. There has been population ageing due to retention of older and higher-quality researchers and a large reduction in the number of younger researchers. Significant differences among universities are evident in the patterns of transformation. The paper also critically considers the PBRF assessment process and characteristics of the metrics used, suggesting scope for improvement in the assessment of researchers and the way in which universities are ranked.

1.   Introduction

The Performance-based Research Fund (PBRF) was introduced in New Zealand in 2003 as a method of allocating research funding to tertiary education organisations on the basis of research performance, rather than the number of students.[1] This is part of a world-wide emergence of performance-based evaluation schemes designed to underpin funding of tertiary institutions; see Coryn (2007), OECD (2010), Jones and Cleere (2014), de Boer et al (2015), and Wilsdon, et al (2015). Similar schemes were introduced earlier in the United Kingdom, Australia and Hong Kong, and subsequently in Denmark, Norway and Sweden and in other countries.[2] These schemes vary by coverage and assessment method, which may be bibliometric or peer-review based. The New Zealand scheme uses a peer-review assessment method and assesses individuals rather than groups.[3]

The Tertiary Education Commission explained the aims of the PBRF process as follows: 'The primary purpose of the Performance-Based Research Fund (PBRF) is to ensure that excellent research in the tertiary education sector is encouraged and rewarded. This entails assessing the research performance of tertiary education organisations (TEOs) and then funding them on the basis of their performance'.[4]

The purpose of this paper is to contribute to understanding the potential effects of these schemes by summarising the results of a research programme analysing the impact on universities of the New Zealand PBRF scheme. The results can contribute towards the wider debate on performance-based schemes for which, as de Boer et al. (2015, p.5) conclude, 'there still is not sufficient evidence on the effects of the systems and ... our

---

[1] For background and detailed discussion of the PBRF, see New Zealand Tertiary Education Commission (2002, 2013) and Ministry of Education (2012).

[2] See also Ministry of Education (2013) for a comparison of the New Zealand scheme with other schemes, and Roberts (2005) on the Australian scheme.

[3] In their summary of the literature regarding research evaluation and metrics, Wilsden, et al (2015, p. viii) argue that, 'Peer review is not perfect, but it is the least worst form of academic governance we have.' An example of how the use of bibliographic metrics can affect research evaluation results in the context of the PBRF is provided by Tressler and Anderson (2012, p. 17) who conclude 'that the time-lags between publication and citation make it difficult to rely on citation counts to produce a meaningful measure of output in a PBRF-like research-evaluation framework, especially one based on the assessment of individual academics'. See Davidson (2013) for a different perspective in the context of the Australian ERA system.

[4] See http://www.tec.govt.nz/funding/funding-and-performance/funding/fund-finder/performance-based-research-fund/.

understanding of the proper design and implementation of performance agreements is still incomplete'.

The PBRF has changed the incentives facing universities and individual researchers. It is important to examine both the nature of those incentives and their potential effects. Improvement in a university's research quality – assessed as some form of average of individual researcher qualities – can come about for three reasons. First, changes arise from recruitment and retention of high-quality staff; second, from encouraging the exit of lower-quality staff; and third, from converting low-quality to high-quality researchers. Each of these paths has different 'effective prices' and faces various constraints. The speed of transformation is affected by the supply of high-quality researchers and the form of university contracts, including the flexibility in pay scales. The effective prices and constraints are likely to differ among universities. It is neither feasible nor desirable to have a university consisting only of researchers judged by the PBRF process to be of the highest quality at a given date: the balance between high-quality and low-quality researchers is also likely to vary between universities. This paper explores the possible implications of PBRF by examining the nature of the relevant flows and their contribution to changes in NZ universities' average research quality.

Others have endeavoured to assess the impact of PBRF by comparing publication measures before and after the introduction of PBRF.[5] However, no previous work has utilised the data on individual researcher performance produced by the PBRF process itself. The insights summarised in this paper are derived using a database consisting of an anonymous 'quality category' for each individual assessed in each of the three PBRF assessment rounds in 2003, 2006 and 2012. The 2006 round was a partial round and hence the results discussed in this paper refer to the changes between 2003 and 2012.[6] This database, which was provided by the Tertiary Education Commission (TEC) under a

---

[5] Examples of this type of work as applied to the economics discipline include Gibson *et al*. (2008), Anderson *et al*. (2013), Anderson and Tressler (2014). For examinations of a wider range of disciplines, see Smart (2009) who concluded that research output and quality increased for most New Zealand universities after the introduction of PBRF; Hodder and Hodder (2010) for New Zealand, and Butler (2003) for Australia, who concluded that research quantity increased but quality declined. A critique of Butler is provided by van den Besselaar *et al*. (2017), who find that both quantity and quality of research has increased, and refer to other studies showing a positive relationship between research quantity and impact.

[6] In 2006, universities could choose to submit a new portfolio for researchers covering the previous six years of research, or retain the quality category of those researchers who submitted portfolios in 2003 covering the six years prior to 2003.

confidentiality agreement and is not publicly available, includes the quality category, age, discipline and university of each individual assessed from 2003 to 2012.

The paper is structured as follows. Key features of the process used to assess researcher quality are explained in Section 2. Section 3 reports changes in the number of researchers between 2003 and 2012. A summary of the change in the quality of researchers at New Zealand universities since the introduction of PBRF is provided in Section 4. This section also discusses the relative performance of each university. Section 5 examines the contribution of exits and entrants, along with transitions among quality categories at the aggregate level. Section 6 evaluates these transitions at the level of each university, revealing how they vary. Although information about management practices within universities is not available, a number of reasons for these different responses are discussed. The associated change in the age-distribution of researchers is discussed in Section 7. The metrics used in the PBRF are critically evaluated in Section 8. Conclusions and policy issues are discussed in Section 9.

2.   Key features of the PBRF process

The PBRF uses three measures to allocate Government funding to support research at universities and other New Zealand TEOs; these are 'Quality Evaluation', 'Research Degree Completions' and 'External Research Income'. The quality categories form the basis of the Quality Evaluation component, which comprises 60 per cent of PBRF funding.[7] Furthermore, the funding per quality category varies across disciplines.

The allocation of the Quality Evaluation component of PBRF funding is based on information submitted by each university researcher which is referred to as an 'evidence portfolio'. The information provided by the evidence portfolios is used to assess the research quality of each researcher. The evidence portfolio is evaluated by relevant subject expert panels to eventually assign each evidence portfolio to a quality category.

The subject panel assigns a score from 0 to 7 for each of three components of the evidence portfolio: 'research output'; 'peer esteem'; and 'contribution to research environment'.[8] These three components are given scores, s_j, which are then given weights, q_j, of 0.70,

---

[7] Since 2016 the percentage allocations have changed to 55:25:20 respectively.
[8] The TEC guidelines refer to these scores as 'unweighted individual component scores'.

4

0.15 and 0.15. The Total Weighted Score, *S*, for an individual is obtained by multiplying the weighted sum of the s_j values by 100. Hence:

$$S = 100 \sum_{j=1}^{3} q_j s_j \tag{1}$$

The maximum individual score is therefore 700. A letter grade, indicating the quality category, is then assigned depending on the assessed total. These are as follows: R for scores 0 to 199; C for scores between 200 and 399; B for scores from 400 to 599; and A for scores from 600 to 700.[9] A numerical score, *G*, is then assigned to each letter grade: 10 for an A; 6 for a B; 2 for a C; and 0 for R.[10] These categories are described further in Ministry of Education (2012, p. 21). When individuals move between institutions, all their quality score goes to the new institution, irrespective of how long they have been at the new university and how much has been invested in them by the previous employer.[11]

3.   Changes in the number of researchers assessed by PBRF

The TEC dataset comprises 7,041 individual staff portfolios in the 2003 PBRF round, 6,865 in the 2006 round, and 6,652 in the 2012 round. Table 1 reports the total number of university staff portfolios per university assessed in the 2003 and 2012 rounds, using the dataset provided by TEC. These numbers are integers since the employment weights (that is, the full-time equivalent employment proportions) are not available. The number of evidence portfolios in the 2012 PBRF round compared to the 2003 round fell by 389, or 5.5 per cent. The size and proportion of changes vary substantially between the universities.

The period since 2003 coincides with changes in policy with respect to former Colleges of Education. Following the 2003 PBRF round, the Auckland, Christchurch, Dunedin and Wellington Colleges of education were subsumed within the University of Auckland, University of Canterbury, University of Otago and Victoria University of Wellington

---

[9] Only research output from the previous six years qualifies for assessment.

[10] The recognition that new researchers may take time to establish their research, publications, and academic reputations led to the introduction in 2006 of new categories, C(NE) and R(NE), although having the same numerical score as C and R respectively.

[11] However, if they moved within a year of the census date, the funding is divided between the two relevant institutions.

respectively.[12] The Colleges of Education at Hamilton and Palmerston North had been amalgamated with University of Waikato and Massey University respectively prior to the 2003 PBRF. These amalgamations in turn triggered reforms of the former colleges and significant reductions in staff, particularly at the Auckland and Wellington colleges. The total number of staff in the Education discipline who submitted PBRF portfolios was 747 in 2003 and 543 in 2012. This fall of 204 submissions between 2003 and 2012 accounts for 52 per cent of the fall of total submissions between 2003 and 2012 observed in Table 1.

Table 1.   PBRF Evidence Portfolios per University, 2003 to 2012

University	2003	2012
AUT	615	466
Lincoln	203	213
Massey	1,297	1,047
Auckland	1,660	1,776
Canterbury	660	668
Otago	1,370	1,314
Waikato	565	514
VUW	671	654
NZ Total	7,041	6,652

Another factor influencing the change in the number of researchers assessed was the change in 2012 to the process. In the 2003 and 2006 PBRF rounds, all eligible staff were assigned a quality category and if an evidence portfolio was not submitted the staff member was assigned a quality category of R.[13] At a late stage in the 2012 process, TEC changed the rules such that only submitted portfolios were assigned a quality category. Hence universities could choose not to submit portfolios for those who were expected to be assigned a quality category of R. This affected the calculation of AQSs, as discussed further below.

---

[12] The staff numbers for the Universities of Auckland, Canterbury, Otago and Victoria University of Wellington used in this paper include the staff from these respective Colleges of Education.

[13] However, in 2006, those who submitted a portfolio in 2003 did not need to submit a new one if they took the view that their quality category had not changed. In those cases, they were assigned their 2003 category.

University	2003		2012	
	N_P	N_T	N_P	N_T
AUT	135.3	892.7	429.5	952.1
Lincoln	139.1	221.1	174.1	250.9
Massey	689.3	1326.8	918.6	1316.2
Auckland	1152.5	1669.7	1556.1	2023.8
Canterbury	497.7	695.1	617.3	661.2
Otago	845	1297.8	1168.2	1567.5
Waikato	369.8	685.7	440.6	601.3
Victoria	459.9	606.4	641.5	779.9
Total	4288.6	7395.2	5945.9	8152.9

The TEC publishes aggregate university data on an employment-weighted basis, shown in Table 2.[14] The published aggregates show an increase in employment-weighted evidence portfolios between 2003 and 2012, which contrasts strongly with the unweighted numbers reported in Table 1. In 2003, in particular, New Zealand universities employed a considerable number of people on a part-time basis, the extent of which varied among universities.

Table 2. Number of FTE Evidence Portfolios (N_P) and Number of Non-Administration Staff (N_T)

University	2003		2012	
	N_P	N_T	N_P	N_T
AUT	135.3	892.7	429.5	952.1
Lincoln	139.1	221.1	174.1	250.9
Massey	689.3	1326.8	918.6	1316.2
Auckland	1152.5	1669.7	1556.1	2023.8
Canterbury	497.7	695.1	617.3	661.2
Otago	845	1297.8	1168.2	1567.5
Waikato	369.8	685.7	440.6	601.3
Victoria	459.9	606.4	641.5	779.9
Total	4288.6	7395.2	5945.9	8152.9

---

[14] The aggregates therefore give the numbers of 'full-time equivalent persons'.

Figure 1 shows that all universities increased their average FTE per portfolio. For example, in 2003 AUT submitted 615 evidence portfolios, but these converted to only 135.3 on a Full-Time Equivalent (FTE) basis; this implies an overall employment weight of only 0.22, which increased to 0.92 by 2012. In 2003, Auckland submitted 1,660 portfolios, which translated to 1152.5 using employment weights; this implies an overall weight of 0.69, which increased to 0.87 by 2012. The ratio for all universities combined increased from 0.61 to 0.89. This general increase, which occurred at the same time as the total number of evidence portfolios declined, reflects the reduction in the proportion of part-time researchers.

Figure 1 Average Full-Time Equivalent Researchers
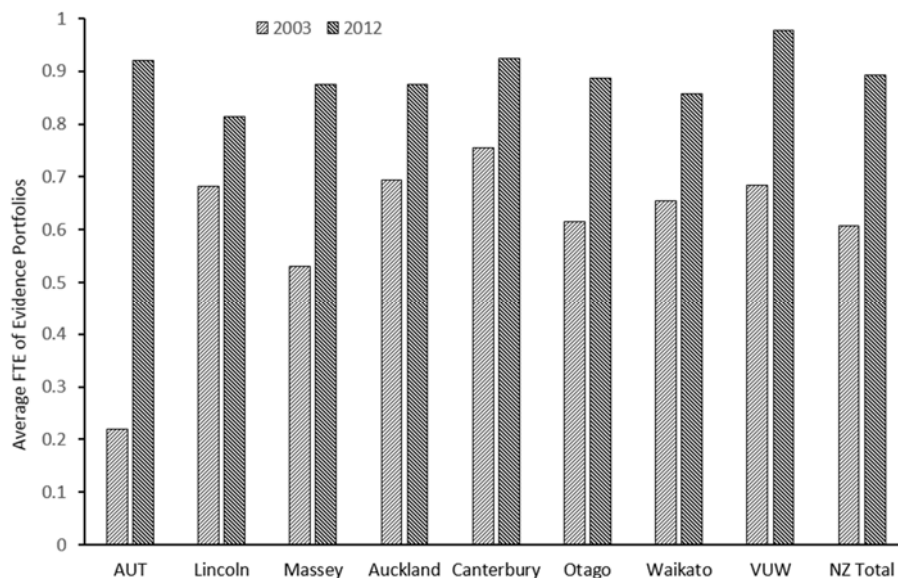
Table 2 also shows the number of employment-weighted evidence portfolios and total non-administration staff for each university in 2003 and 2012. The corresponding ratios are shown in Figure 2. The ratio for the entire university system increased from 0.58 in 2003 to 0.73 in 2012. The ratios varied substantially in 2003. AUT's ratio in 2003 was only 0.15 while VUW's ratio was 0.76. All universities increased this ratio by 2012, and particularly AUT for which the increase was three-fold. The rise in the ratio of employment-weighted portfolios to total employment-weighted non-administration staff

implies that universities reduced the proportion of non-administration staff who were not research active.[15]

Figure 2 Ratio of FTE Evidence Portfolios to Total Number of Staff
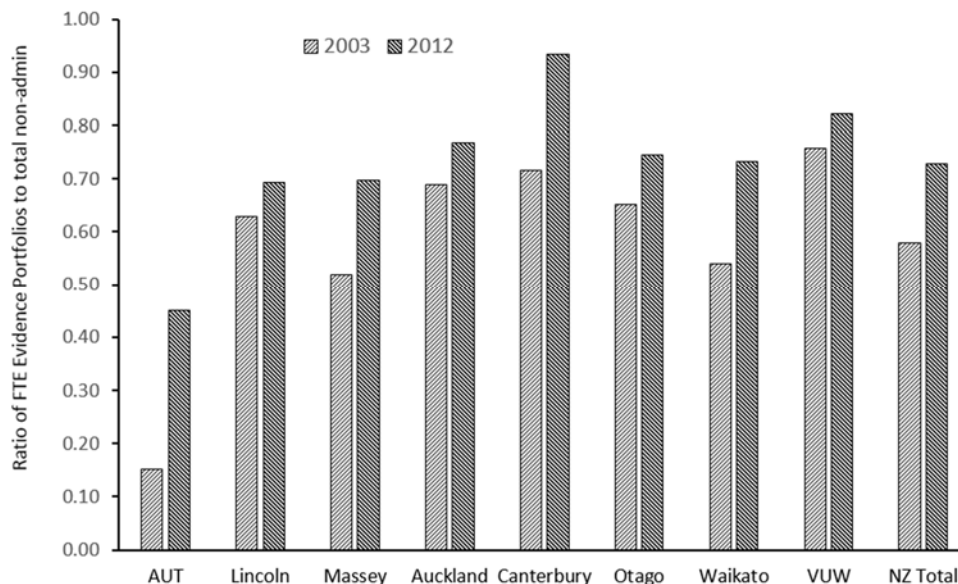
4. Changes in average quality scores for New Zealand universities

This section reports changes between 2003 to 2012 in the average quality scores, AQSs, of the eight universities, derived as follows. Define the employment weight of person i as $e_i \leq 1,$ and let n denote the relevant number of employees in a university.[16] The choice of n has an important effect and is discussed in more detail below. The AQS is:

$$AQS = \sum_{i=1}^{n} e_i G_i \bigg/ \sum_{i=1}^{n} e_i \qquad (2)$$

As explained in Section 2, the value for each researcher, G, ranges from zero to 10, depending on the individual's quality category. Hence, a university's AQS can range from zero to 10.[17]

---

[15] Using as a measure of research active those individuals publishing at least one article, in whole or in part, during a 6-year period in an EconLit listed journal, Anderson and Tressler (2014, p. 7) concluded that the proportion of research active academic staff increased from 71% in 1994–9 to 79% in 2000–5 and 83% in the period 2006–11.

[16] As mentioned above, when using the anonymous TEC dataset, the employment weights are all set to unity.

[17] There is an inconsistency in the terminology used by TEC. The term 'Total Weighted *Score*' refers, for each individual, to the weighted sum (multiplied by 100) of the scores given for the three components

The Tertiary Education Commission produced a range of AQS measures, depending on the choice of denominator. One measure uses the employment-weighted total of all researchers for whom PBRF portfolios were submitted, while another measure excludes R-type researchers. Since the grade for R-type staff is equal to zero, their inclusion affects only the denominator in (2). The change in the rules in 2012, which allowed universities to withhold submission of the portfolios of potential R-type researchers, means that the AQSs published by the Tertiary Education Commission for 2012 are not necessarily comparable with those for 2003.[18]

However, an alternative preferred measure can be derived by using the total number of non-administration staff as the denominator in equation (2). The use of this denominator avoids the problems of comparability mentioned in the previous paragraph. An AQS based on all non-administrative staff can be obtained by multiplying the AQS which includes all portfolios, including R-researchers and based on the unweighted TEC data, by the ratio of the employment-weighted number of portfolios to the total number of non-administrative staff. The latter ratio is obtained from the published data in Table 2. The details of this calculation are explained in Buckle and Creedy (2018).[19]

Table 3 shows AQSs derived on this basis for each university and all universities combined, and indicates that all universities have improved.[20] However, there are substantial differences in the rates of improvement across the eight universities. The proportional growth rates for each university are shown in Table 4. The final row of this table shows the extent to which the growth in each AQS exceeds that which would arise in a scoring system under which all non-R researchers are given the same score. It can be seen that, for all universities combined, the change in the distribution of C, B and A researchers has contributed only an additional 9 per cent to the growth of the AQS. This

---

of research quality. The term 'Average Quality *Score*', is not in fact an average of *scores,* but is a weighted average of the *values* attached to individual quality categories, where the latter are determined by the individual's position in the distribution of Total Weighted Scores.

[18] Two additional measures include as the denominator either the sum of effective full-time students or the sum of postgraduate students. These are affected by the discipline mixture and in particular by student-staff ratios, and hence do not necessarily reflect average research quality of staff.

[19] The AQS values published by the Tertiary Education Commission use employment-weighted data for both numerator and denominator, but they do not report AQSs based on all non-administrative staff.

[20] This increase complements the finding of Gemmell *et al*. (2017) that research productivity in NZ universities has increased markedly since the early 2000s.

appears to suggest that there has been little overall improvement over and above the removal of a large number of non-research-active staff.

Table 3 Average Quality Scores Based on All Non-Administration Staff: 2003 to 2012

University	2003	2012
AUT	0.110	1.364
Lincoln	1.579	2.458
Massey	1.071	2.878
Auckland	2.457	3.736
Canterbury	2.532	4.262
Otago	2.002	3.555
Waikato	1.579	3.145
VUW	2.310	4.374
All	1.733	3.307

However, there has been considerable variation among universities. For example, in VUW the change in the quality composition of researchers (arising both from recruitments and transformations) has produced a growth in its AQS that is 83 per cent higher than it would be under a scoring system that gives all non-R researchers the same weight. The equivalent contribution for Lincoln was 52 per cent, and yet this university had the second-lowest growth rate in its AQS. The percentage additional contribution arising from the composition of researchers is lowest for AUT, Waikato and Canterbury, yet they are in the group having the highest growth rates in their AQSs. Hence, no clear pattern of change emerges.

Table 4. Growth of AQSs from 2003 to 2012

	AUT	Mass	Linc.	Wai.	Otago	VUW	Auck.	Cant.	NZ
AQS 2003	0.11	1.07	1.58	1.58	2.00	2.30	2.46	2.53	1.73
AQS 2012	1.36	2.88	2.46	3.15	3.55	4.37	3.74	4.26	3.31
% growth	1143	169	56	99	78	89	52	68	91
% Improvement	Percentage improvement of AQS growth rate above that which would result from a common score (G) for A, B and C Quality Categories								
	10	29	52	14	29	83	25	17	9

The question arises of whether the changes exhibit convergence, implying an inverse relationship between the proportional improvement of AQSs and initial AQSs. Table 4 arranges universities from left to right from the lowest to highest AQS in 2003. Convergence is consistent with growth rates decreasing from left to right. AUT has the

lowest AQS in 2003 and has by far the highest growth rate of over 1,100 per cent. Massey has the second lowest AQS in 2003 and the second highest growth rate at nearly 170 per cent. However, the remaining observations do not follow the pattern predicted by convergence: there is no systematic tendency towards uniformity of AQSs across universities.[21]

A related issue concerns the policy objectives of the PBRF scheme. While one stated purpose is to raise the quality of research in NZ universities, this can create a tension between 'equity' and 'efficiency' in research funding; see, for example, Hicks and Katz (2011). The incentives may have the effect of generating a concentration of higher-quality research in a small number of universities. However, objections may be made on equity grounds involving for example, the geographical distribution of researchers.

5.   Patterns of change in research quality of NZ universities

This section examines the characteristics of research quality change within NZ universities. This is achieved by considering the range of flows within universities which contributed to changes in their AQSs. First, transitions among the quality categories, along with entrants and exits from 2003 to 2012, for all universities combined are shown in Table 5. The flows are from rows to columns: for example, of those who were As in 2003, 54.8 per cent were As in 2012 in the same university.

These flow percentages show, for example, that a very small proportion, just under 6 per cent, of those who enter (that is, submit a portfolio for their employing university in 2012 for the first time) do so with a grade R in 2012. The largest exit rate, of just over 70 per cent, is for those who were classed as R in 2003.[22] Hence, low recruitment and high exit rates of Rs is a strong feature of transitions over the PBRF period.

A low proportion, 7.5 per cent, of entrants between 2003 and 2012 were classified as A-researchers in 2012. The majority of entrants are classed as B and C in 2012, at 35 and 52 per cent respectively. Just under 20 per cent of the 2003 Rs moved upwards to become Cs in the same institution by 2012, and 6 per cent moved upward to B. Upward

---

[21] More formal testing of convergence of AQSs is provided in Buckle and Creedy (2017a).
[22] However, some Rs may have entered another university at the same or a higher level. And, as discussed above, some could have been given new contracts which meant that they avoided the need to submit a portfolio.

movements within the same institution came mainly from C and B researchers, where for each category about a quarter of individuals moved one step upwards from 2003 to 2012. Hence for Bs and Cs, it could be said that about one third of entrants (between 2003 and 2012) fell into each category, and about a quarter of those who were classed as either B or C in 2003 progressed to a higher grade in the same institution by 2012.[23]

Table 5. Matrix of Flows (Percentages): All Universities Combined 2003 to 2012

Quality category in 2003	Quality Category in 2012					
	A	B	C	R	Exits	Total
A	54.8	16.1	0.5		28.7	442
B	18.1	38.4	8.4		34.1	1772
C	2.9	26.3	23.3	1.2	46.4	2374
R		6.0	19.2	4.6	70.2	2453
Entrants	7.5	34.7	51.8	5.9		3165
Total	868	2622	2833	329	3554	10,206

It also seems that of those classed as C-researchers in 2003, a high proportion, 46 per cent, had exited by 2012. It is likely that many of these moves were to another university, and may have involved some kind of promotion, but this information is not available from the data. The B-researchers also experienced substantial outward mobility, and 29 per cent of those who were A in 2003 had exited. This is consistent with a small proportion of entrants over the period being classed as A-researchers in 2012. There can be a large proportion of As who left between 2003 and 2012, since the denominator (initial number of As) is much smaller than the total number of entrants in all categories over the period.

6.   Individual university responses

This section examines the extent of heterogeneity among universities in their recruitment and transitions. These results reported in Section 5 suggest that the conversion of Bs and Cs to higher grades is less 'costly' than that of Rs. The relative ability to recruit higher quality researchers and retain high quality researchers may vary across university.

---

[23] It has been noted that in 2006, a distinction was made between 'new and emerging' (NE) for C and R staff. The effect of the NE category on upward movement was examined using matrices of transitions between 2006 and 2012. The distinction in the case of Cs appears to have had little value. But in the case of R researchers, those who were NE in 2006 experienced much more upward movement than other Rs. This suggest that the PBRF encouraged more careful selection of entry-level researchers.

Universities were incentivised to recruit at the higher levels and appoint very few researchers with low research outputs. However, the incentives vary among universities depending on their initial AQS. For example, it can be shown that the proportional increase in the AQS of a university with n researchers (assuming these are all full-time employees), resulting from hiring one additional type-A researcher, is given by:

$$\text{Proportional change in AQS} = \frac{\left(\dfrac{w_A}{AQS} - 1\right)}{n+1} \qquad (3)$$

Where w_A is the value assigned to a quality category of A. By contrast, the change in AQS from eliminating n_R type-R people is given by:

$$\text{Proportional change in AQS} = \frac{n_R}{n - n_R} \qquad (4)$$

Figure 4 illustrates some variations in the change in AQS for the case where $n = 700$. Clearly, if a university has an AQS greater than 2, its AQS would in fact fall as a result of hiring one more C-type researcher. It would need to recruit at least at level B, or reduce the number of R-type people. There are sharply decreasing returns, in terms of the impact on AQS, from hiring A and B researchers. These properties suggest that university responses to the introduction of PBRF are likely to vary depending on their initial situation.

To illustrate the varying responses, Table 6 provides information about the flows of researchers from 2003 to 2012 for each university. The actual flows are shown on the left-hand side. It is also possible to construct hypothetical flows, on the assumption that each university has the same transition proportions as all universities combined, but start from their actual numbers in 2003. The right-hand side of Table 6 shows the differences between the actual flows and the hypothetical flows. If all entries in this right-hand block were zero, all universities could be said to display identical transition propensities. A formal test of whether a university differs from 'All NZ universities' can be obtained by computing a standard chi-square statistic, using the hypothetical flows as 'expected'

values. The resulting values are reported in each case on the same line as the university name in Table 5.[24]

Figure 3 Effects on AQS of Varying Quality Categories
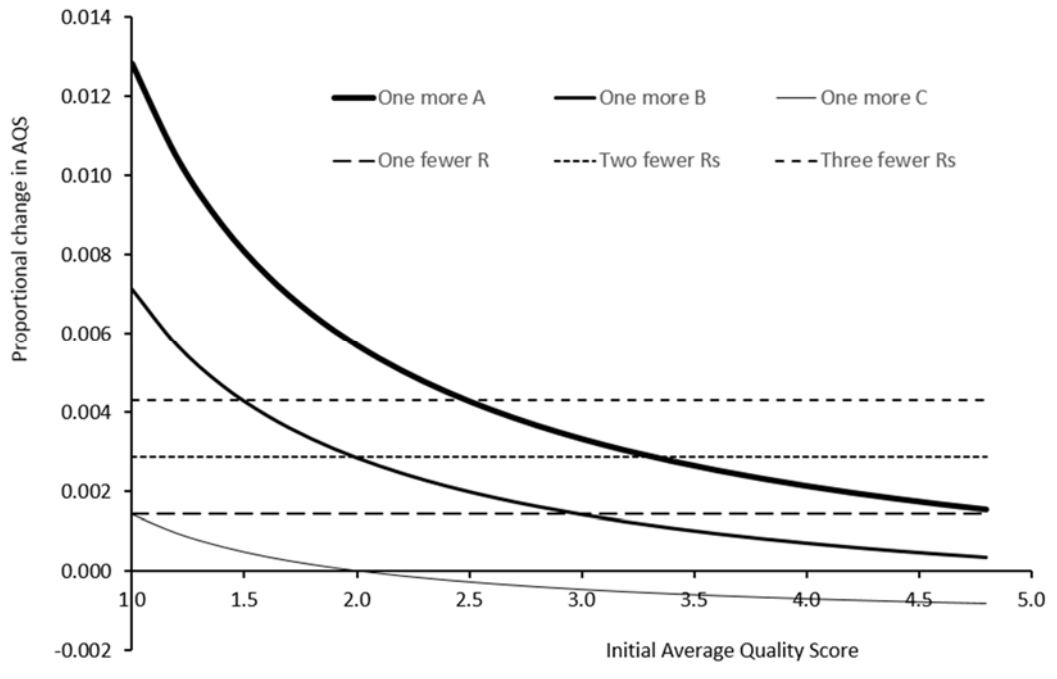
As anticipated, the transition proportions clearly differ among universities. The only universities whose transitions do not differ significantly from those of all universities combined are Massey and Auckland. The university that differs most from the average is VUW, which had the highest AQS in 2012. The next-highest chi-square value is for AUT, which had the highest percentage change in its AQS over the period. While AUT and VUW differ most from the overall pattern, they also differ from each other considerably. For example, if VUW is compared with AUT (that is, by computing expected frequencies by combining VUW transition and exit rates with the initial AUT stocks) a chi-square value of 323.6 is obtained.[25]

---

[24] The appropriate chi-square values, for type I errors of 0.05 and 0.10, and for 16 degrees of freedom, are 26.296 and 23.542 respectively.

[25] This is the highest value in such pairwise comparisons: for example, comparing VUW with Auckland and Lincoln give chi-square values of 250.7 and 245.6 respectively. The lower-ranked universities also differ significantly from each other: hence comparing Lincoln with AUT and Massey give chi-square values of 54.51 and 174.5 respectively.

Table 6. Actual Flows and Differences Between Actual and Hypothetical Flows for Each University: 2003 to 2012.

	Actual					Differences from actual				
	A	B	C	R	Exits	A	B	C	R	Exits
AUT	*Chi-square =73.75*									
A	3	2				0	+1	0	0	-1
B	3	18	1		9	-2	+6	-2	0	-2
C	2	29	24	2	48	-1	+1	0	+1	-1
R		11	84	22	357	0	-17	-7	0	+24
Entrants	9	52	171	33		-11	-40	+34	+17	
Lincoln	*Chi-square =50.54*									
A	5	1			1	+1	0	0	0	-1
B	7	17	6		11	-1	+1	+2	0	-3
C	2	22	30	2	39	-1	-3	+8	+1	-5
R		0	10	9	41	0	-4	-2	+6	-1
Entrants	5	18	64	15		-3	-17	+11	+9	
Massey	*Chi-square =17.67*									
A	25	6			13	+1	-1	0	0	0
B	34	83	27		79	-5	0	+7	0	+3
C	14	101	117	6	212	+1	-17	+12	+1	+3
R		34	93	23	430	0	-1	-19	-4	-23
Entrants	27	157	273	27		-9	-11	+22	-2	
Auckland	*Chi-square =16.74*									
A	93	17	1		48	+6	-9	0	0	+2
B	115	192	45		186	+16	-19	-6	0	+2
C	20	153	115	8	253	+4	+9	-13	+1	-2
R		29	86	20	279	0	+4	+6	+1	-11
Entrants	80	303	444	55		+14	-3	-13	+3	0
Canterbury	*Chi-square =59.06*									
A	24	9	1		21	-6	0	+1	0	+5
B	31	78	9		95	-3	+5	-9	0	+22
C	5	59	49	0	140	-2	-8	-10	-3	+23
R		12	47	10	70	0	+4	+20	+4	-28
Entrants	20	124	173	17		-5	+8	0	-3	
Otago	*Chi-square =37.03*									
A	56	16			24	+3	+1	0	0	-4
B	78	161	43		101	+2	0	+4	0	-30
C	14	127	130	1	207	0	+1	+19	-5	-15
R		26	67	14	305	0	+1	-12	-5	+16
Entrants	43	211	313	14		-1	+9	+12	-20	
Waikato	*Chi-square =57.33*									
A	12	13			9	-7	+8	0	0	-1
B	18	59	25		52	-9	+1	+11	0	-1
C	5	61	54	6	69	-1	+10	+9	+4	-21
R		9	51	13	109	0	-2	+16	+5	-19
Entrants	12	76	84	16		-2	+11	-13	+5	
VUW	*Chi-square =92.05*									
A	24	7			11	+1	0	0	0	-1
B	34	73	10		72	+2	+6	-6	0	+7
C	6	72	33	4	133	-1	+7	-25	+1	+18
R		26	34	2	130	0	+14	-3	-7	-5
Entrants	42	158	119	10		+17	+44	-52	-9	

Inspection of Table 6 reveals that the universities differ from each other in different ways.[26] As suggested by Figure 3, changes evident in the universities with a lower initial AQS differ from those with a higher initial AQS. Thus AUT, which had the lowest initial AQS, recruited relatively few A and B researchers and converted few people to higher grades. It was relatively strong in recruiting Cs and achieving a high number of exits from Rs, but it also recruited more Rs than average. Lincoln also recruited relatively more Cs and Rs, but was about average in its ability to convert researchers to higher grades and achieving exits of Rs. Massey also recruited relatively more Cs, was quite successful at converting Rs to Cs, and in achieving exits of Rs.

Thus, the experience differed among the three universities that were consistently ranked at the low end of the AQSs. The common features for this group are the low recruitment of As and Bs, and their reliance on recruiting Rs, with improvements coming from recruiting relatively more Cs and getting high exits among Rs. Conversion rates were also low, with the exception of Massey regarding Rs. It could be argued that these universities face relatively high effective costs of recruiting As and Bs. Nevertheless, these universities managed to retain relatively more of their higher-scoring researchers.

Auckland was strong at recruiting As, but were in other respects similar to the pattern revealed by all universities combined. Canterbury recruited relatively more Bs, but at the same time suffered more than average exits of A and B researchers. It was also successful at converting Rs to Cs. Otago recruited relatively few Rs and had higher exits of Rs, while being strong at recruiting Bs. Nevertheless, it was about average at converting researchers to higher-scoring academics. Waikato had a relatively strong conversion of Rs to Cs and were stronger at appointing Bs, while achieving fewer exits of Rs.

At the top of the AQS ranking in 2012, VUW achieved success at appointing As and Bs, and keeping its higher-scoring staff. In addition, it appointed relatively fewer Cs and Rs. Unusually, it had success in converting Rs to B researchers. Among the top five universities, Canterbury stands out in terms of its ability to appoint from people who were rated B researchers in 2012. Canterbury's higher losses of As and Bs may be explained to some extent by the earthquake experience, since those researchers are expected to be

---

[26] In comparing performance relating to the exits of R-researchers, it has to be kept in mind that this figure is distorted for reasons discussed in the introduction and in Section 4.

more mobile. The higher-ranked universities were also relatively better at converting researchers to higher grades, with the possible exception of Otago.

It might be argued that these higher-ranked universities have the kind of environment, including stronger academic leadership from A and B researchers, and perhaps also more resources devoted to internal research grants, which stimulate higher outputs. Also, in recruiting Rs, they may also be selecting from those who are at the higher end of the scale in terms of their potential. The greater ease of attracting As and Bs to the higher AQS universities, combined with the effect of the funding formula, may be said to involve an effective reduction in their 'price'.

One difficulty in interpreting the findings is that the analysis of transitions during which PBRF has operated does not have a control group with which to compare performance.[27] Hence, it is not obvious that all the changes can be attributed directly to the PBRF process. Nevertheless, there is no doubt that New Zealand universities in 2003 had a large number of Rs, and this was inconsistent with the incentive structure created by the PBRF exercise. Indeed, the high exit rate of those categorised as R in 2003, combined with the low recruitment rate over the 2003 to 2012 period of those categorised R in 2012, combined with the relatively high recruitment of B and C researchers, is most unlikely to represent an equilibrium situation.

This question can be formally assessed by considering the implied equilibrium distribution generated by the transitions and entries observed over the PBRF period, if they are held constant. After a sufficient time, the system would settle into an equilibrium distribution of academics across the quality categories: in this situation, outward movements from each category would be balanced by inward movements each period. An equilibrium distribution derived by using the transition matrix and vector of births in Table 5 was found to give equilibrium stocks of each quality category that are not feasible and very different from the totals observed for 2012.[28] It is therefore not unreasonable to

---

[27] Ideally, in order to isolate the effect of PBRF it would be necessary to compare the post-PBRF transitions with those that would have occurred in the absence of PBRF. This would require information about previous transitions, including the likely behaviour of those who may not have left the system in the absence of PBRF. In addition, there may have been other external influences, such as the effect of the global financial crisis on the ability of New Zealand universities to recruit academics from other countries, the effects of which cannot be isolated.

[28] See Buckle and Creedy (2018) for further details.

suggest that the evolution of New Zealand universities since 2003 represents to some extent a structural shift in response to the introduction of PBRF, and this path cannot be expected to continue indefinitely. These results help to provide an indication of the extent to which the changes since 2003 are in fact large, despite the many characteristics of universities which make structural change difficult.

7. Changes in the age distribution of researchers

The desire to retain A and B researchers, who are typically relatively older, along with considerable caution in recruiting entry-level researchers, so that only those who have already displayed some publishing success are likely to be appointed to lectureships, suggests a tendency for population ageing in the NZ university system. The age of some researchers was not reported in the TEC dataset, and hence the sample size used to evaluate changes in the age distribution of researchers is smaller than that used in the other Sections. Nevertheless, Table 7 shows a marked change in the age distribution of researchers. Between 2003 and 2012, there were large declines in the number of researchers in each age group, except for the 60-89 group, with a 77.7 per cent growth rate. The largest absolute decline was in the 40-49 age group, which fell by 415 researchers (a decline of 18.2 per cent). But the largest percentage decline was in the 20–29 years age cohort which fell by 64.3 per cent.

Table 7. Change in the Age Distribution of PBRF Researchers From 2003 to 2012

| | Age | | | | | |
	20-29	30-39	40-49	50-59	60-89	Totals
Absolute change in number in age group	-153	-244	-415	-231	+602	-441
Change as per cent of 2003 age group	-64.3	-16.1	-18.2	-10.9	+77.7	-6.4

The changes shown in Table 7 may be associated with changes in the age distribution by grade. To compare the distributions, the observed proportions for 2003 were used to derive 'expected frequencies' for the 2012 age distribution by grade and 2012 grade distribution by age (using the absolute totals in 2012). The differences between the observed and expected frequencies for the 2012 age distribution by grade are shown in Table 8: these may be referred to as 'unexpected changes'.

The sum of each column in Table 8 shows the differences for each age group between observed and expected frequencies; the sum of each row is zero. These totals reveal that the large increase in the frequency of researchers in the 60-89 age group is much higher (by 599.71) than could be expected from the 2003 marginal frequencies. Conversely, the decline in the frequency of researchers in each of the other age groups is larger than would be expected from the 2003 marginal frequencies. The largest differences are the decline of approximately 306 in the 40-49 age cohort, and the rise of approximately 600 older age cohort researchers.

Table 8. Differences Between Observed and Expected 2012 Age Distribution by Quality Category

QC	Age				
	20-29	30-39	40-49	50-59	60-89
A	0.00	-13.10	1.77	-79.90	91.23
B	3.22	-82.09	-130.50	-77.55	286.92
C	-29.32	2.27	-157.85	-3.43	188.33
R	-15.34	-13.75	-19.25	15.11	33.23
Totals	-41.44	-106.67	-305.83	-145.77	599.71

However, there were higher than expected numbers of R-category people in the older age groups. Conversely, there are large unexpected declines in the number of people in C and R categories aged 20-29, of Bs in the 30-39 age cohort, of C and B researchers aged 40-49, and large unexpected declines in A and B researchers aged 50-59.

The unexpected high number of older people of R-quality may be because it is more difficult and expensive to remove people in the older age cohorts.[29] The 'excess' decline in lower-category researchers in the youngest age cohort may reflect the possibility that people in that stage of their careers are more career mobile or are easier to remove. [30]

8. Assessment of PBRF appraisal method and metrics

Having examined some of the characteristic of changes in AQSs, it is useful to return to the choice of metric itself. A fundamental initial judgement required in any research quality appraisal involves the question of whether quality is perceived to be a continuous

---

[29] In some cases, they may be valued for administration or teaching qualities.

[30] Formal chi-square tests showed that the differences in Table 7 statistically significant at all reasonable significance levels.

(or quasi-continuous) variable, or whether it is discrete, whereby individuals are placed into a small number of well-defined categories. The basic view of 'quality' has not been articulated by the Ministry of Education and the PBRF process involves what might be described as a hybrid (or perhaps ambivalent) method.

As explained in Section 2, each research portfolio is assigned an initial total weighted score by discipline experts, which is the sum of the weighted individual component scores for research output, peer esteem and contribution to research environment. These initial scores are subject to a further round of assessment, where adjustments may be made. After this second round, the portfolio is then placed into one of four quality categories, R, C, B, and A.

The method used to derive total weighted scores, involving equation (1), has some properties which may not be immediately apparent. With three components s_j, where $j = 1, 2, 3,$ each is given an integer score from zero to seven, the total weighted score, S, can range from 0 to 700, but many values in the range are not possible. The number of possible values actually depends on the weights, q_j, used to form the aggregate, S. The number of permutations is 512. With PBRF weights of 0.7, 0.15 and 0.15, the possible values start from 0 and increase in increments of 15 until reaching 60. The next two numbers in the sequence are 70 and 75, after which they increase in increments of 5 up to 570. The next number is 580 and then the values increase in increments of 5 up to the maximum of 700. Furthermore, the number of ways of achieving a given value of S varies with S itself. To give a few examples, there are two ways of obtaining a score of 15, but 5 ways of getting a score of 60, 6 ways of obtaining 345, 7 ways of getting a score of 90, and 8 ways of getting 315. This means that, if everyone has an equal independent probability of obtaining a score from 0 to 700 for each of the quality types, there is not an equal probability of falling into equal ranges of S.

Buckle and Creedy (2017b) show that the scoring system can itself have a substantial effect on the measured quality distribution of researchers. Examination of the distribution of initial weighted scores in the 2003 and 2012 PBRF rounds, and movements in those scores during the expert panel assessments stages, indicates that the use of quality categories and associated thresholds appear to have influenced the performance scores assigned to research portfolios and the final allocation of portfolios to quality categories.

In particular, the process generated large spikes in the distribution of total weighted scores at the lower thresholds for each quality category. Furthermore, the distribution of scores is roughly symmetric. These features contrast strongly with the distribution of metrics used by other performance evaluation methods. Without a clear rationale for the use of quality categories and the selection of threshold values, the choice of weights, q_j, assigned to each component also appears arbitrary. Yet, these weights and the distribution of total weighted scores can affect the research quality ranking of universities and disciplines.

This raises questions as to whether the presence of threshold levels created by the use of quality categories, and the weights attached to each component, influence the decisions of peer reviewers and discipline expert panels in assigning preliminary scores and moving people to the next threshold. It also raises the possibility that the final distribution of total weighted scores has properties that differ from those which would emerge in a system using only an overall measure based on the distribution of weighted scores, S, for disciplines and universities, rather than the AQSs based on the weighted quality categories.

This system used by the Tertiary Education Commission to summarise research quality discards information about the relative quality of researcher portfolios which could be used to derive alternative summary measures of average university quality.[31] Furthermore, it is not possible to know what distribution of scores would have arisen without the reviewers' prior knowledge of the use of quality category thresholds. An alternative method, using quality categories, would involve the independent evaluation of scores, followed by the determination of thresholds. This is still subject to the criticism that the use of quality categories compresses all scores within a given range to a single value.

Considerable attention is given to the AQSs and resulting rankings, in view of the kind of reputation effects stressed by OECD (2010). Nevertheless, as mentioned in the introduction, funding is based on a formulae which uses individual quality categories, rather than AQSs, with funding weights that vary by discipline.

---

[31] The variation could perhaps be regarded as 'noise' rather than a clear quality signal. If this view is taken it is not clear why the first stage is used, as individuals could be assigned directly to discrete categories.

9. Conclusions

This paper has summarised the changes in NZ universities' measured research quality since the introduction of the PBRF scheme in 2003, which was designed to promote and reward research excellence in the tertiary education sector. The New Zealand scheme is particularly interesting because it uses a peer review system of assessment and assesses all eligible university staff.

The 'Quality Evaluation' component of the PBRF funding formula takes account of the number of full-time equivalent researchers and their quality. It was found that both these components have increased, despite a reduction in the number of evidence portfolios submitted by universities. Furthermore, all universities increased their submitted evidence portfolios as a ratio of their total number of non-administrative staff.

Based on the data for individual quality categories, the average quality of researchers has nearly doubled between 2003 and 2012. This inference is based on a measure of the average quality score (AQS) for all universities combined, where the AQS is the ratio of the sum of all quality categories expressed as a proportion of the sum of all university non-administration staff. Although the TEC has published other AQS measures, they either do not capture the average quality of researchers, or they are compromised by the decision in 2012 by TEC to remove R-quality researchers from the calculations.[32]

There are considerable differences in the rates of improvement amongst the eight universities. The fastest rates of improvement were achieved by AUT and Massey universities, two of the lower-ranked universities in terms of AQSs in 2003. The slowest rate of improvement was experienced by Lincoln, also one of the lowest ranked in 2003. Rates of improvement also varied markedly amongst the top-ranked universities. Consequently, there has not been a clear tendency for convergence to the mean level of research quality for the eight universities. There has been persistence in the relative quality of the group of top-ranked and the group of bottom-ranked universities. Rankings have changed within the groups, with Auckland declining amongst the group of higher-

---

[32] Researchers using bibliometric measures to evaluate the impact of PBRF on research quality have drawn varying conclusions. For example, using data from the Web of Science, Smart (2009) concludes that research output and quality increased at New Zealand universities since the introduction of PBRF. However, Hodder and Hodder (2010) conclude that quantity increased while quality decreased.

ranked universities and Lincoln declining in the lower group, but there has been a persistence in membership of the two quality groups.

A strong feature contributing to improvements in research quality was a large reduction in R-quality staff. In addition to low recruitment and high exit rates of R-quality researchers, it also seems that a high proportion of those classed as C-researchers in 2003 had left their 2003 university by 2012. The majority of entrants were B and C quality researchers. Upward movements within the same university came mainly from C and B researchers. But the experiences varied significantly across universities, and this variation tended to be associated with the research ranking of the university in 2003. In particular, of the new entrants between 2003 and 2012, the higher-ranked universities tended to be more successful at attracting and retaining A and B quality researchers. This was a standout feature of Victoria which experienced the highest proportional improvement in its AQS compared to the other higher 2003 ranked universities of Auckland, Canterbury and Otago. The lower-ranked universities in 2003, consisting of AUT, Massey and Lincoln were more successful at recruiting C-quality researchers.

It is not clear whether the PBRF process was intended to encourage a more concentrated allocation of funding to the higher research-quality universities. While there is not strong evidence for convergence or divergence in quality growth rates, the responses have been found to depend on each university's initial research quality. Thus, the characteristics of growth for AUT and Massey were very different from, for example, VUW and Canterbury. Furthermore, despite the high growth rates of AUT in particular and, to a lesser extent, Massey University, the margin in the AQSs between the top and bottom ranked universities has persisted. It would be helpful if the Ministry of Education were to make an explicit statement about whether is believes that a more efficient use of funding would involve a greater concentration of research quality or, on the other hand, it prefers a more dispersed allocation. The preferred goal should influence the design of the performance-based funding schemes and their incentives.

Care needs to be taken in judging the AQSs. For all universities combined the AQS is below 4, out of a maximum of 10. Because of its unique features, it is not possible to compare the New Zealand universities with those elsewhere. Whether this value is judged to be poor or satisfactory really depends on a clear view of what is being measured and

what optimal balance of staff is appropriate, since a situation where most non-administrative members of a university are A-quality may not be desirable. The Ministry of Education has not expressed a view on what it regards as a 'satisfactory' AQS for the university system. This is very important when considering the merit or otherwise of retaining the scheme and its associated metrics.

A strong feature associated with the improvements in researcher quality has been a significant change in the age distribution of researchers. There was substantial population ageing over the period. This arose from a combination of an increase in the average age of entry and reduced exits from older age groups. This led to changes in the age distribution within quality category grades. The cost of recruiting younger researchers who do not progress to higher quality categories has increased with the introduction of PBRF. This incentive has indeed influenced recruitment, with a tendency to recruit mainly early career academics with clearer evidence of research capability. However, there has not been a relatively high rate of exit from older Rs, and this may be because of employment related constraints associated with tenure.

Without similar pre-PBRF data, it is not obvious that the observed changes can all be attributed to the introduction of the PBRF process. Nevertheless, it can be argued that the transformation that has taken place is not sustainable, in the sense that if the transition proportions remained constant over time, the equilibrium distributions of researchers would be unrealistic. It is therefore reasonable to suggest that the large changes over the period 2003 to 2012 have, to a substantial degree, been stimulated by the introduction of the PBRF and that such large changes are unlikely to be repeated.

The PBRF scoring system involves several stages of evaluation and weighting to derive each individual's quality category. This combines the use of an idiosyncratic quasi-continuous scoring system with the use of discrete quality categories. It reflects an ambivalence concerning the view taken of how research quality should be measured. The system could be significantly simplified given a clear judgment on this crucial issue. For example, if it is believed that research quality can only be classified into discrete broad categories, there would seem to be no role for a first stage derivation of a total weighted score for each person. On the other hand, if continuous variation over a wide quality range, reflected in many metrics of research quality, is believed to be appropriate, the

PBRF scoring system could be considerably simplified and improved to reflect the kind of heterogeneity revealed in other metrics.

Where the PBRF process is used to rank universities in terms of research quality, there are no good reasons not to include, as denominator in the calculation of AQSs, the total number of non-administrative staff.[33] This is the denominator used in the present paper, although it has not been used by the Tertiary Education Commission (which excludes R-quality researchers) when publishing PBRF outcomes. In addition, the use of student numbers as one denominator has no clear rationale and the publication of AQSs on this basis has been subject to misinterpretation.

It is clear that an incentive structure involving some kind of assessment is needed to allocate public funds and to stimulate the adoption of appropriate policies by universities. A challenge is to avoid unintended negative consequences. This paper has shown that the research quality of NZ universities has improved substantially and it can be argued that this is a direct result of the introduction of the PBRF system. Nevertheless, it cannot be concluded unambiguously that universities have been appropriately rewarded for their improvements or that the procedures and metrics used are ideal. Given the large compliance costs, there is clearly scope to improve the system in respect of the demands on universities and academics in producing their evidence portfolios, the way in which these are evaluated, and the way funds are allocated. The debate on reforms to PBRF would also need to address issues raised by other commentators regarding, for example, the high administrative and compliance costs, the types of research supported, and incentives to work on contemporary NZ issues.

References

Anderson, D.L. and Tressler, J. (2014) The New Zealand performance-based research fund and its impact on publication activity in economics. *Research Evaluation*, 23, 1-11.

---

[33] Indeed, a case may be made for including all staff.

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*, 47, 140-157.

Buckle, R.A. and Creedy, J. (2017a) The evolution of research quality in New Zealand Universities as measured by the Performance-Based Research Fund process. *Victoria University of Wellington Working Papers in Public Finance*, no. WP11/2017.

Buckle, R.A. and Creedy, J. (2017b) An evaluation of metrics used by the Performance-Based Research Fund Process in New Zealand. *Victoria University of Wellington Working Papers in Public Finance*, no. WP16/2017.

Buckle RA, Creedy J (2018) The Evolution of Research Quality in New Zealand Universities as Measured by the Performance-Based Research Fund Process. *New Zealand Economic Papers* (forthcoming). Available online at: http://www.tandfonline.com/doi/full/10.1080/00779954.2018.1429486

Butler, L. (2003) Explaining Australia's increased share of ISI publications -The effects of a funding formula based on publication counts. *Research Policy*, 32, 143–155.

Coryn, C.L.S. (2007) *Evaluation of Researchers and Their Research: Toward Making the Implicit Explicit.* PhD Thesis, Western Michigan University, USA.

Davidson, S. (2013) Excellence in research for Australia: an audit of the Applied Economics Rankings. *Agenda*, 20, 5-20.

de Boer, H. Jongbloed, B. Benneworth, P. Cremonini, L. Kolster, R. Kottmann, A. Lemmens-Krug, K. and Vossenstey, H. (2015) Performance-based funding and performanceagreementsinfourteenhighereducationsystems, Report for the Ministry of Education, Culture and Science, Centre for Higher Education Policy Studies, C15HdB014, Universiteit Twente, Enschede. www.utwente.nl/cheps

Gemmell, N., Nolan, P. and Scobie, G. (2017) Estimating quality-adjusted productivity in tertiary education: methods and evidence for New Zealand. *Victoria University of Wellington Working Papers in Public Finance*, no. WP17/2017.

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. *University of Waikato Department of Economics Working Paper*, 08\11.

Hicks, D. and Katz, J.S. (2011) Equity and excellence in research funding. *Minerva*, 49, pp. 137-151.

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.

Jones, A. and Cleere, L. (2014) Furthering the research impact of UCD: report of the Beyond Publications Committee. University College Dublin. Available at: http://hdl.handle.net/10197/7292

Ministry of Education (2012) A History and Overview of the PBRF. Wellington: Ministry of Education. Available at: https://education.govt.nz/assets/Documents/Further-education/Policies-andstrategies/Performance-based-research-fund/PBRFHistoryAndOverview.pdf

Ministry of Education. (2013) *An international comparison of performance-based research funding schemes.* Retrieved from http://www.minedu.govt.nz/NZEducation/EducationPolicies/TertiaryEducation/PolicyandStrategy/~/media/MinEdu/Files/EducationSectors/TertairyEducation/PBRF/InternationalComparison.pdf

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

New Zealand Tertiary Education Commission (2013) *Performance-Based Research Fund: Evaluating Research Excellence – the 2012 Assessment*. Wellington: Tertiary Education Commission. (Available from Evaluating Research Excellence - the 2012 Assessment).

OECD (2010) Performance-Based Funding for Public Research in Tertiary Education Institutions: Workshop Proceedings. OECD Publishing. Available from: http://www.oecdilibrary.org/education/performance-based-funding-for-publicresearch-intertiary-education-institutions_9789264094611-en

Roberts, G. (Chair) (2005) Research quality framework: assessing the quality and impact of research in Australia. Department of Education, Science and Training Expert Advisory Group for a Research Quality Framework: Issues Paper. Available at: www.csu.edu.au/__data/assets/pdf_file/0019/51472/rqf_issuespaper.pdf

Smart, W. (2009) *The Impact of the Performance-Based Research Fund on the Research Productivity of New Zealand Universities*. Wellington: Ministry of Education. Available from: www.educationcounts.govt.nz

Tressler, J. and Anderson, D. L. (2012) Citations as a measure of the research outputs of New Zealand's economics departments: the problem of 'long and variable lags'. *Agenda*, 19, 17-34.

van den Besselaar, P., Heyman, U. and Sandström, U. (2017) Perverse effects of output-based research funding? Butler's Australian case revisited. *Journal of Informetrics*, http://dx.doi.org/10.1016/j.joi.2017.05.016

Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., Jones, R., Kain, R., Kerridge, S., Thelwall, M., Tinkler, J., Viney, I., Wouters, P., Hill, J. and Johnson, B. (2015) The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management. (DOI: 10.13140/RG.2.1.4929.1363).

About the Authors

Robert Buckle is Professor Emeritus at Victoria Business School,
Victoria University of Wellington
Email: bob.buckle@vuw.ac.nz

John Creedy is Professor of Public Finance at Victoria Business School,
Victoria University of Wellington
Email: john.creedy@vuw.ac.nz

VICTORIA
UNIVERSITY OF WELLINGTON
TE WHARE WĀNANGA
O TE ŪPOKO O TE IKA A MĀUI

Chair in Public Finance
Victoria Business School

Working Papers in Public Finance