# The Evolution of Research Quality in New Zealand Universities as Measured by the Performance-Based Research Fund Process

Robert A. Buckle and John Creedy

WORKING PAPER 11/2017
July 2017

**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 Working Papers in Public Finance** series is published by the Victoria Business School to disseminate initial research on public finance topics, from economists, accountants, finance, law and tax specialists, to a wider audience. Any opinions and views expressed in these papers are those of the author(s). They should not be attributed to Victoria University of Wellington or the sponsors of the Chair in Public Finance.

Further enquiries to:
The Administrator
Chair in Public Finance
Victoria University of Wellington
PO Box 600
Wellington 6041
New Zealand

Phone: +64-4-463-9656
Email: cpf-info@vuw.ac.nz

Papers in the series can be downloaded from the following website:
http://www.victoria.ac.nz/cpf/working-papers

Working Papers in
Public Finance

# The Evolution of Research Quality in New Zealand Universities as Measured by the Performance-Based Research Fund Process[1]

Robert A. Buckle and John Creedy

Victoria University of Wellington

**NOTE: This paper corrects (September 2018) some numbers reported in the earlier version posted in July 2017 (in Tables 2, 9 to 12)**

**Abstract**

This paper examines how the research quality of staff within New Zealand universities has evolved since the introduction in 2003 of the Performance-based Research Fund (PBRF). The analysis uses a database consisting of an anonymous 'quality evaluation category' (QEC) for each individual assessed in each of the three PBRF assessment rounds. Emphasis is on the evaluation of organisational changes in performance. The paper examines the extent to which each university's Average Quality Score (AQS) changed as a result of changes in the QECs of existing staff over time and from the exit and entry of staff with different scores. The sensitivity of university rankings to the cardinal scale used by the PBRF is also considered and the degree of convergence amongst the universities is assessed. The data also include information about the age of staff evaluated in PBRF, and this is used to evaluate changes in the age distribution of staff across universities, and the ages of those making transitions within universities and between grades. The results reveal a systematic ageing of university staff in NZ and a significant change in the grade distribution by age, and age distribution by grade. A number of hypotheses regarding organisation change in response to the introduction of PBRF are discussed and tested by comparing universities with different patterns of change.

**Key words:** education, New Zealand universities, Performance-based Research Fund, productivity, research.

**JEL classifications:** I2; I23; I28.

[1] We are grateful to Amber Flynn and Leon Bakker for assisting engagement with New Zealand Tertiary Education Commission (TEC), and TEC for providing the data. We have benefited from discussions with Sharon Beattie, Shelly Biswell, Jonathan Boston, Norman Gemmell, Gary Hawke, Adam Jaffe, Julie Keenan, John MacCormick, Katy Miller and Grant Scobie. An earlier version of this paper was presented at the New Zealand Association of Economists Conference, Wellington, July 2017.

## 1. Introduction

The purpose of this paper is to evaluate how the research quality of staff within New Zealand's eight universities, as measured by the Performance-based Research Fund (PBRF) process, has changed since the scheme was introduced in 2003. The analysis is based on observations of the anonymous 'Quality Evaluation Category' (QEC) for each individual staff member assessed in each of the three PBRF assessment rounds, undertaken in 2003, 2006 and 2012.

The PBRF scheme was designed to unbundle the research component of Government funding of New Zealand tertiary education organisations (TEOs) and allocate the funding on the basis of research performance rather than on the basis of the numbers of students enrolled. The Tertiary Education Commission explained the aims 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'.[2] Similar performance-based funding schemes were introduced earlier in the United Kingdom, Australia and Hong Kong, and subsequently in Denmark, Norway and Sweden. These schemes vary by coverage and assessment methods, which may be bibliometric or peer-review based; see Ministry of Education (2013, p. 4). The New Zealand scheme uses a peer-review assessment method and assesses individuals rather than groups. This differs from the otherwise similar Research Assessment Exercise (RAE) in the United Kingdom.[3]

Universities participating in the PBRF process, faced with incentives in terms of funding and reputation, can attempt to change incentives facing staff and potential graduate students.[4] For example, they can make internal research grants available, change their recruitment policies to target higher-calibre researchers, and change the funding

---

[2] See http://www.tec.govt.nz/funding/funding-and-performance/funding/fund-finder/performance-based-research-fund/. See also Mahoney (2004), and Smart and Engler (2013).

[3] Since 2014, the UK RAE scheme has been known as the Research Excellence Framework (REF). On other schemes, see, for example, Bakker *et al.* (2006) and Ministry of Education (2013). On the Australian exercise, see Roberts (2005). For a summary of international approaches, see Coryn (2007), Jones and Cleere (2014).

[4] There are variations in the extent to which universities depend on public funds; this may affect the strength of incentives created by PBRF.

allocations to departments. Furthermore, they can change tenure rules, performance management procedures and retirement incentives, increase scholarship funding to attract more graduate students, and so on. The funding incentives also extend to differential funding for Maori and Pasifika researchers. However, it may be that universities are heavily constrained in their ability to make substantial structural changes.

Systematic evidence about actual management practices in universities and how they responded to the introduction of the PBRF scheme is not available. However, the changes over time in the QEC of anonymous individuals who submit a research portfolio and remain in the same institution in the different PBRF rounds, along with information about exits and entrances to those institutions, can be examined to obtain information about the characteristics of change within organisations. Using the QECs allocated by expert review panels to each university staff member, it is possible to see how the average quality score for each university has evolved by examining the transformation of the make-up of eligible staff at each university. That is, the extent to which a university's average quality score changed as a result of changes in the quality scores of existing staff and from the exit and entry of staff with different QECs can be determined.

The challenge facing the TEC and designers of the PBRF process was to create the right incentive structure to foster the appropriate changes and, importantly, to reveal the necessary information. The measurement process and metrics used to assess the quality of research for each person must also be widely accepted as reasonable: this is extremely difficult to achieve when those making the assessments do not typically have direct familiarity with the research and when substantial differences among academic disciplines exist.

It must be recognised that any research appraisal process creates an environment in which a certain amount of gaming, or strategic and opportunistic behaviour, is likely to take place. This includes lobbying, for example regarding the selection of individuals to sit on assessment panels, and applying pressure to change the rules. One aspect of the PBRF process is that it has appeared to create an adverse incentive regarding the inclusion of those for whom a research portfolio was submitted. Instead of requiring information relating to all non-administrative staff, there are several criteria relating to the definition of teaching and research staff who fall under the definition of 'eligible for PBRF' and for whom a portfolio must be submitted. For example, staff whose contracts indicate that they are under supervision, or are less than 0.2 of a full-time equivalent, or whose contracts ended before the submission date, were not required to submit a portfolio. The manipulation of contracts is of course a part of the variety of responses to PBRF, and was highlighted by the auditors; see KPMG (2013).

Section 2 describes the PBRF data used in the present analysis. Section 3 considers the number of researchers assessed in each university in each of the three PBRF rounds. The changes in average quality scores for each university over the period 2003 to 2012 and derived from the QEC for each researcher, are examined in Section 4. This section also considers the sensitivity of university rankings to the cardinal scale used by the PBRF; alternative average quality scores, relating to the choice of denominator in the averaging process, are also reported. The question of whether there has been convergence among the universities is also examined in Section 4. Section 5 examines overall changes in the age and grade distributions. Several hypotheses regarding organisation change in response to the introduction of PBRF are suggested in Section 6. These are tested in Section 7 using information about transitions. Section 8 considers the sustainability of these transitions by producing equilibrium distributions. Brief conclusions are in Section 9.

It may be worth alluding to a number of aspects of the PBRF process which cannot be examined here. For example, despite the change to the funding formula away from student numbers to an emphasis on research performance measures, it may be argued that universities have long attached a high priority to research, and individual promotions and reputations have depended on recognised outputs. The introduction of PBRF has involved a shift in priorities and, by attaching cardinal scores to different outputs, has changed the way outputs are judged within universities. This has given more strength to those attempting to push for certain changes; see Walsh (2004).

Other commentators have suggested that perverse incentives, and unintended consequences, may have been created by PBRF. These include compromising academic freedom, encouraging researchers to favour conventional over more innovative research agendas, a shift away from teaching, a focus on shorter-term projects (ruling out large-scale scholarly works), and less attention being given to New Zealand policy issues (given a perceived greater need to publish in overseas journals). On these issues, see, for example, Curtis (2004), Curtis and Matthewman (2005), Ministry of Education (2008), Whitley (2008), OECD (2010), Clear and Clear (2012), and Butler and Mulgan (2013). Another criticism made of research assessment schemes is they can discourage publication in second-tier journals which serve important training and knowledge bridging roles, and publish topics not well served by mainstream journals; see Chavarro, Tang and Ràfols (2016).

The PBRF process also involves high costs, and a full evaluation would also consider the wider costs and benefits of the exercise, as argued by Hazledine and Kurniawan (2005).[5] The benefits of research are not quantified and an optimal level and composition of funding is not considered. The aim of PBRF was to allocate total research funding among universities in order to maximise the aggregate research output or performance, and this aim is considered separately from the question of the optimum total research investment.[6]

A further characteristic of a process like the PBRF is that there is inevitably a learning process by university managers and individuals about the way in which evidence is used to make assessments. It is likely that some of the increase in average quality scores represents this costly learning process rather than a genuine quality improvement. However, it is not possible to separate these components. Furthermore, drawing inferences about the improvement in research quality over the three assessment rounds requires an assumption that the approaches used by successive (and changing) expert panels did not change significantly.[7]

## 2. The nature of the PBRF quality assessment

There are three measures used to allocate Government funding to support research at

[5] A Ministry of Education (2013a, pp. 7-8) report found that, 'The total estimated transaction costs for universities and the Tertiary Education Commission associated with the Quality Evaluation for the six-year period between 2007/08 and 2012/13 was $52.1 million. This amounts to just under 4% of the PBRF funding allocated in that period. Universities bore the majority of these costs, which were driven primarily by time spent by staff compiling and assessing evidence portfolios'.

[6] The introduction of PBRF has created considerable interest in the relative performance of universities and disciplines, and improvements in the PBRF process: see for example, Boston *et al*. (2005), Boston (2007).

[7] There is in fact a complex process of calibration and moderation undertaken by the Tertiary Education Commission during each stage of assessing portfolios in order to encourage consist assessment over time and across discipline panels. This process includes training expert panel members on assessment methods and referring to a sample of previous evaluations.

universities and other Tertiary Education Organisations (TEOs). The first component is 'Quality Evaluation' which, during the first three rounds of the PBRF, comprised 60 per cent of the Fund allocated on the basis of an assessment of the research quality of eligible staff. The second component is 'Research Degree Completions', which comprised 25 per cent. The third component is 'External Research Income', which comprised 15 per cent and is based on the amount of external research revenue generated.[8]

The PBRF has therefore sharpened incentives for universities to improve the research calibre of staff, to increase research degree completions and to increase funding for research from alternative external sources. The higher the average research rating of a TEO's eligible staff members, for a given number of staff, the higher is the proportion of funds allocated to that particular TEO. Similarly, the higher the number of eligible research active staff at a TEO for a given quality and number of eligible staff at other TEOs, the higher the PBRF funding allocated to that particular TEO. Furthermore, the incentive to improve the research calibre of staff may also have been enhanced by the 'Research Degree Completion' and 'External Research Income' components of the PBRF.

The Tertiary Education Commission (TEC) publishes an 'Average Quality Score' (AQS) for each university at the conclusion of each PBRF round. To evaluate each researcher, each PBRF portfolio is assessed by a panel assigned to a discipline or group of disciplines. Only the previous six years are considered, and it does not necessarily consider all relevant information since the portfolio can mention only a limited number of publications and other contributions to the research environment.

The relevant subject panel assesses the quality of each portfolio and assigns a score from 0 to 7 for each of three categories: these are 'research output'; 'peer esteem'; and 'contribution to research environment'. These three scores, $s_i$, are given weights, $w_i$, of 0.70, 0.15 and 0.15. The total score, $S$, for an individual is obtained by multiplying the weighted sum of the $s_i$ values by 100. Hence:

$$S = 100\sum_{j=1}^{3} w_j s_j \qquad (1)$$

[8] Since 2016 the percentage allocations have changed to 55:25:20 respectively.

Thus the maximum individual score is 700. A letter grade, referred to by the TEC as the 'quality evaluation category', QEC, is then assigned depending on the assessed total. These are as follows: R or R(NE) for scores 0 to 199; C or C(NE) for scores between 200 and 399; B for scores from 400 to 599; and A for scores from 600 to 700. An A is intended to reflect 'international standing'; B reflects 'national standing' and C reflects 'local standing'. An R denotes that an individual is not considered to be research active. The recognition that new researchers may take time to establish their research, publications, and academic reputations led to the introduction in 2006 of the new categories, C(NE) and R(NE). These categories applied to new and emerging researchers who did not have the benefit of a full six-year period.

A numerical score, $G$, is then assigned to each letter grade: 10 for an A; 6 for a B; 2 for a C or C (NE); and 0 for R and R (NE). A university's average quality score, AQS, is the employment-weighted mean score, which can range from zero to 10. Define the employment weight of person $i$ as $e_i \leq 1$ , and let $n$ denote the relevant number of employees in a university. The choice of $n$ has an important effect and is discussed in more detail below. The average quality score is:

$$AQS = \frac{\sum_{i=1}^{n} e_i G_i}{\sum_{i=1}^{n} e_i} \tag{2}$$

Since the grade for R-type staff is equal to zero, their number affects only the denominator in (2). The sensitivity to this choice is considered in subsection 4.4.[9]

The funding per grade varies across disciplines. An institution's total funding, based on the Quality Evaluation component of PBRF, is therefore determined by the sum of the funding attracted by each assessed researcher, which in turn is determined by the sum of the product of individual quality evaluation categories and the discipline funding per individual score.

[9] TEC have also produced a range of measures of AQS which vary according to the choice of denominator, including the sum of effective full-time students and 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.

The scoring system therefore involves the assignment of three cardinal scores, $s_i$, to obtain an aggregate score, *S*, for each individual, and then a further conversion of *S* to *G*, before the AQS for each institution is computed. However, with an initial numerical scoring system (out of 700) there is really no need to convert to A, B, C and then convert each of those to a single number. This clearly raises the question of why an intermediate conversion, involving considerable compression within wide ranges of the *S* scores, is used. In other words, this homogenises individuals within each of the categories (A, B, C, R). The rationale for this transformation is not clear.[10] A university with many people just below the cut-off points is judged, perhaps unfairly, to be much lower in quality than one with many people who just get above the thresholds.

The dataset provided by TEC gives staff quality grade letters used to determine an institution's average quality score. The TEC dataset comprises 7,052 individual staff portfolios in the 2003 PBRF round, 6,879 in the 2006 round, and 6,664 in the 2012 round. Only the individual QECs are available: the separate components discussed above are not provided. Each individual's age and discipline are also available.

The three PBRF rounds were conducted in 2003, 2006, and 2012. An important late change was made in 2012 regarding the assessment process. Calculation of the average quality score for each university in 2003 and 2006 included those graded R. As mentioned above, the inclusion of Rs affects the denominator of equation (2) but not the numerator, since Rs are given a score of zero. Hence the exclusion of Rs in 2012 meant that the AQS values for those universities with relatively more Rs increased by more than those with few Rs, compared with the AQS they would have achieved if Rs had been included.

Another feature of PBRF is that 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.

An important qualification is that pre-PBRF data on research quality are not available. Hence, it is not possible to compare the dynamics during the PBRF period with earlier periods. However, it is possible to examine how post-PBRF dynamics contributed to improved performance. Furthermore, comparisons among universities can provide

[10] It may perhaps have been influenced by a desire to report results easily in a simple table of the number of As, Bs and so on, rather than, say, using summary measures of the distribution of individual scores.

information about the nature of successful (that is, AQS-improving) organisational change.

## 3. The number of researchers assessed in each PBRF round

In view of the changes made both to the PBRF evaluation process and the criteria determining whether a portfolio is submitted, it is necessary to consider several alternative measures of the 'total number of staff'. Let $N_P$ denote the number of people who submit a portfolio, while $N_R$ denotes the number of R-graded portfolios, and finally $N_{NP}$ denotes the number for whom there is no portfolio. Then $N_T = N_P + N_{NP}$ is the total number of non-administrative staff employed by the university.

Information about the values of $N_T$ and $N_P$ can be obtained from Tertiary Education Commission (2013), which gives weighted totals using the appropriate part time fraction. These are shown in Table 1. The table reveals substantial variations in the ratio, $N_P / N_T$, over time and among universities. The lowest value was for AUT in 2003 (0.152) and the highest was for Canterbury in 2012 (0.934). The ratio rose consistently over the period for AUT, Massey, Canterbury, Otago and Waikato, and fell between 2003 and 2006 for Auckland and VUW.

***Table 1 Ratio of Funded Evidence Portfolios ( $N_P$ ) to Number of Staff ( $N_T$ )***

| University | 2003 | | | 2006 | | | 2012 | | |
|---|---|---|---|---|---|---|---|---|---|
| | $N_P$ | $N_T$ | $N_P/N_T$ | $N_P$ | $N_T$ | $N_P/N_T$ | $N_P$ | $N_T$ | $N_P/N_T$ |
| AUT | 135.3 | 892.7 | 0.152 | 221.8 | 890.2 | 0.249 | 429.5 | 952.1 | 0.451 |
| Lincoln | 139.1 | 221.1 | 0.629 | 165.9 | 221.9 | 0.748 | 174.1 | 250.9 | 0.694 |
| Massey | 689.3 | 1326.8 | 0.520 | 873.8 | 1396.2 | 0.626 | 918.6 | 1316.2 | 0.698 |
| Auckland | 1152.5 | 1669.7 | 0.690 | 1241.2 | 1859.7 | 0.667 | 1556.1 | 2023.8 | 0.769 |
| Canterbury | 497.7 | 695.1 | 0.716 | 549.7 | 692.1 | 0.794 | 617.3 | 661.2 | 0.934 |
| Otago | 845.0 | 1297.8 | 0.651 | 990.0 | 1401.9 | 0.704 | 1168.2 | 1567.5 | 0.745 |
| Waikato | 369.8 | 685.7 | 0.539 | 417.3 | 599.8 | 0.696 | 440.6 | 601.3 | 0.733 |
| Victoria | 459.9 | 606.4 | 0.758 | 598.5 | 988.2 | 0.606 | 641.5 | 779.9 | 0.823 |
| Total | 4288.6 | 7395.2 | 0.580 | 5058.4 | 8049.9 | 0.628 | 5945.9 | 8152.9 | 0.729 |

Table 2 reports the total number of university staff assessed in each round, using the dataset provided by TEC. These numbers are therefore integers, since the employment weights are not available. The number of portfolios in the 2012 PBRF round compared to the 2003 round fell by 389, or 5.5 per cent. The size and proportion of change varies substantially between the universities. The differences between the tables reflect the changing distribution of part-time researchers in different universities.

***Table 2 Staff PBRF portfolios ( $N_P$ ) per university, 2003 to 2012***

| University | Number of staff PBRF portfolios submitted: $N_P$ | | | | |
|---|---|---|---|---|---|
| | 2003 | Change | 2006 | Change | 2012 |
| AUT | 615 | -220 | 395 | +71 | 466 |
| Lincoln | 203 | + 21 | 224 | -11 | 213 |
| Massey | 1,297 | -139 | 1,158 | -111 | 1,047 |
| Auckland | 1,660 | +113 | 1,773 | +3 | 1,776 |
| Canterbury | 660 | +39 | 699 | -31 | 668 |
| Otago | 1,370 | -112 | 1,258 | +56 | 1,314 |
| Waikato | 565 | -23 | 542 | -28 | 514 |
| VUW | 671 | +145 | 816 | -162 | 654 |
| NZ Total | 7,041 | -176 | 6,865 | -213 | 6,652 |

Changes in the number of portfolios submitted in each PBRF round result from a complex combination of factors, including strategic variations in the nature of some contracts, mentioned earlier, and changes in the PBRF process itself. Also important are university management decisions in response to enrolment changes, changes in staff management practices, and changes in staff recruitment priorities. Changes in the number of researchers submitted by each university that are observed in the TEC dataset reflect the net effect of: (i) staff who remain with the same university during each round; (ii) staff exiting the New Zealand university system; (iii) recruitment of staff from outside the New Zealand university system (that is, recruitment from elsewhere in New Zealand or from another country); (iv) changes in the categorization of staff within each university (for example, some staff move from academic positions to management positions); and (v) transfers of staff from one university in New Zealand to another.

It is beyond the scope of this study to examine the contribution of all these management practices to the change in the number of research staff portfolios submitted in each PBRF round. However, the period since 2003 coincides with changes in Government tertiary education policy with respect to former Colleges of Education, and there were significant changes in Education faculty staffing levels during this period that are worth mentioning.

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 respectively.[11] 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 of education and significant reductions in staff, particularly at the Auckland and Wellington colleges. The total number of staff in Education 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 2. Associated with these staff reductions were large changes in the grade and age distributions, characteristic of the changes for all disciplines discussed in Section 5 below.

## 4. Changes in university average quality scores

This section analyses changes in the average quality scores in each of the three PBRF rounds at the eight New Zealand universities. Subsection 4.1 compares the official AQS published by the TEC with the AQSs estimated from the TEC database of anonymous individual QECs used in this study. The TEC-reported AQSs are based on the employment weighted QEC for each individual researcher (that is, they use the full-time equivalent weights). The data used in this study do not include employment weights and therefore the estimated AQSs are based on unweighted scores for each individual researcher. Unsurprisingly, there are differences between the TEC reported AQSs and the estimated AQSs derived in this study, but the differences are small and it is reassuring that the ranking of university AQSs derived from the unweighted individual scores are similar to the TEC rankings.

The cardinal score, $G$, assigned by the TEC to each letter grade, could potentially influence the ranking of university AQSs. The reason for the choice of these numerical scores is not clear. Alternative scoring schemes that assign different weights to the letter grades (A, B, and so on) result in different AQSs, and in principle could result in different rankings of university AQSs if the distribution of grades varies between universities. Subsection 4.2 considers the impact on the ranking of university AQSs of alternative

[11] 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.

scoring schemes assigned to the grades for every individual researcher, including the TEC scoring scheme.

Subsection 4.3 presents a preliminary analysis of the source of the changes in each university's AQS. It concentrates on the aggregate components by analysing the contribution of changes in the proportion of R-quality (or non-research active) staff relative to the contribution of changes in the proportion of research-active staff to changes in AQSs for each university. Further comparisons are provided in subsection 4.4, which uses the total number of staff as denominator in the AQS calculation. Finally, subsection 4.5 considers whether the data reveal any convergence in growth rates.

### 4.1 University AQSs and rankings for each PBRF round

Considerable attention has been attached to the average quality scores published by TEC because they provide an indication of the research quality of staff at each university and they can influence a university's reputation.[12] Table 3 shows the official published TEC AQSs for the eight New Zealand universities during each round, weighted by their employment status (full-time or part-time). In 2012 the TEC determined, on the eve of the submission date, that non-research-active eligible staff (those deemed to be equivalent to an R-quality researcher) would be excluded from the calculation of the official AQSs. Therefore, the TEC-published AQSs for 2012 do not include researchers with QEC scores of zero and are not strictly comparable to the AQSs for 2003 and 2006. Table 3 also shows the AQSs derived from the database used in this study, unweighted by employment status.

Table 3 reveals that weighting QECs by employment status does make a difference to the estimated AQS for universities and the university sector, but the differences are small and range from 0.01 (or 0.3 per cent of the TEC AQS for Massey in 2006) to 0.46 (or 12 per cent of the TEC AQS for VUW in 2006). The differences in 2012 are generally much smaller. In contrast to the 2003 and 2006 AQSs, the 2012 AQSs exclude researchers with zero scores (Rs and R(NE)s). If R and R(NE) researchers were more likely to be part-time employers, the differences between the weighted TEC average quality scores and the unweighted average quality scores would be smaller than if they were full-time researchers. The somewhat larger differences between the TEC AQSs and the AQSs

[12] The role of reputation effects is stressed in OECD (2010).

derived from the unweighted scores in 2003 and 2006 may also be attributable to the employment status of R-quality researchers, and the submission decisions of universities. For instance, universities which had relatively more R-quality researchers in 2003 and 2006 would have larger differences between the TEC and estimated unweighted scores reported here. The use of unweighted employment PBRF scores to generate university AQSs results in only a few changes in the ranking of university AQSs compared with the TEC rankings (one change in 2003, two in 2006 and none in 2012).

***Table 3 Weighted and unweighted PBRF Average Quality Scores***

| Uni | 2003 | | | | 2006 | | | | 2012 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | TEC | UW | \|Diff\| | Rk | TEC | UW | \|Diff\| | Rk | TEC | UW | \|Diff\| | Rk |
| AUT | 0.77 | 0.72 | 0.05 | 8 | 1.86 | 1.76 | 0.10 | 8 | 3.59 | 3.44 | 0.15 | 8 |
| L | 2.56 | 2.51 | 0.05 | 6 | 2.97 | 2.92 | 0.05 | 7 | 4.02 | 4.05 | 0.03 | 7 |
| M | 2.11 | 2.06 | 0.05 | 7 | 3.05 | 3.06 | 0.01 | 6 | 4.31 | 4.36 | 0.05 | 6 |
| A | 3.96 | 3.56 | 0.40 | 1 | 4.19 | 3.77 | 0.42 | 2 to 3 | 5.12 | 5.10 | 0.02 | 2 |
| C | 3.83 | 3.54 | 0.29 | 2 | 4.10 | 3.79 | 0.31 | 3 to 2 | 4.79 | 4.76 | 0.03 | 4 |
| O | 3.23 | 3.08 | 0.15 | 4 to 3 | 4.23 | 4.14 | 0.09 | 1 | 4.96 | 4.88 | 0.08 | 3 |
| W | 2.98 | 2.93 | 0.05 | 5 | 3.73 | 3.72 | 0.01 | 5 to 4 | 4.53 | 4.61 | 0.08 | 5 |
| VUW | 3.39 | 3.05 | 0.34 | 3 to 4 | 3.83 | 3.37 | 0.46 | 4 to 5 | 5.51 | 5.45 | 0.06 | 1 |
| Total | 2.59 | 2.81 | 0.22 | | 2.96 | 3.52 | 0.44 | | 4.66 | 4.76 | 0.10 | |

Notes: TEC is the Tertiary Education Commission published AQS which is weighted by employment status; UW is the AQS derived from the individual researcher scores used in this study which are not weighted by employment status; |Diff| is the absolute difference between the weighted and unweighted AQS; Rk is the rank AQS score for each university and shows whether the rank differs between the TEC and UW rank scores. AQSs for 2012 were calculated by excluding R researchers. Using Spearman's rank correlation test, $\rho = 0.936$ for 2003, 0.952 for 2006, and obviously 1.000 for 2012. These all exceed the critical value for 6 degrees of freedom of 0.886, at 2.5 per cent level implying that the null hypothesis of no monotonic association between the TEC and UW rankings is rejected for each PBRF year.

The PBRF rankings differ from the international university rankings. For example, those produced by Quacquarelli Symonds (QS) include 'academic reputation' based on a global survey, and citations per faculty, along with the ratio of faculty to students, employer reputation, the proportion of international students and the proportion of international faculty.[13] The QS ranking is therefore influenced by the age and experience of faculty, university size, discipline mix, and factors which influence preferences of international students. In 2012 the QS ranking was, in order from first to last: Auckland, Otago, Canterbury, VUW, Massey, Waikato, Lincoln, AUT. These ranks are significantly different from the PBRF-based ranks.[14] However, virtually all the difference arises from

[13] For details, see http://www.iu.qs.com/university-rankings/world-university-rankings/.

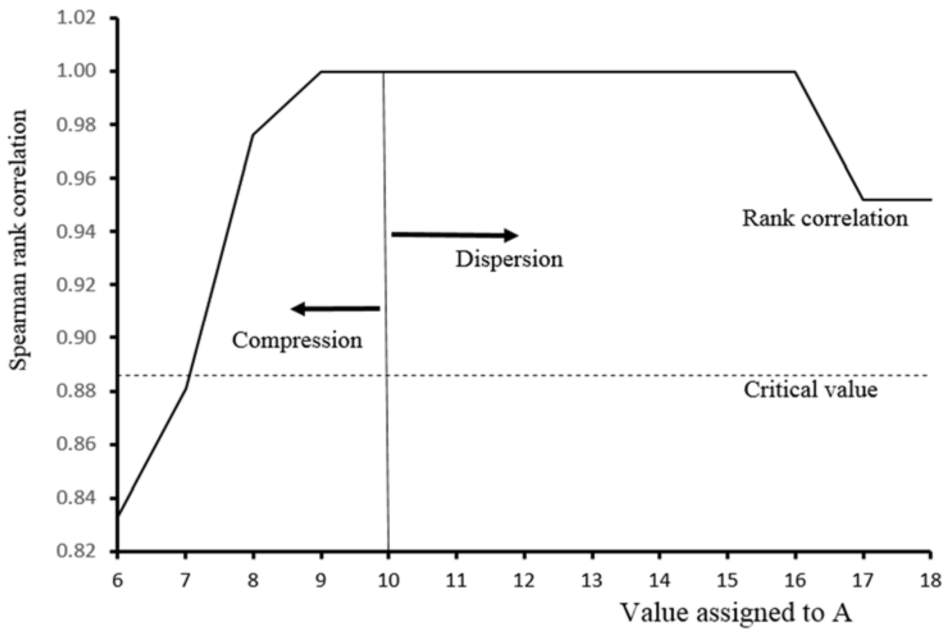
[14] Using Spearman's rank correlation test, $\rho = 0.833$, which is less than the critical value for 6 degrees of freedom of 0.886, at 2.5 per cent level. The null hypothesis of no monotonic association between the QS ranking and the 2012 PBRF ranking of New Zealand universities can therefore not be rejected.

the position of VUW. The lower QS rank for VUW could perhaps arise from other factors which outweigh any gains in research-related quality. For example, the 2011 Christchurch earthquake may have reduced the number of students at Canterbury and Lincoln, and the associated decline in their student/staff ratios helped boost their QS ranking.[15]

### 4.2 Impact of alternative Quality Evaluation Category weights on *AQS*s

The level and ranking of university AQSs may be expected to be sensitive to the choice of numerical score (*G*) assigned by TEC to each QEC. The sensitivity of the AQS levels and rankings can be examined by simulating the impact on each university AQS of alternative numerical scores assigned to each QEC and comparing the results with the outcomes using the TEC weights. For 2012, experiments were carried out by systematically varying the TEC scores (A = 10, B = 6, C = 2, and R = 0), which sum to 18, while leaving the sum of scores unchanged. These experiments involved both compression and dispersion. For compression, this was achieved by systematically reducing the score for As and redistributing to C, and for dispersion the score for As was systematically increased by redistributing from the lower categories, starting from C. In each case a Spearman rank correlation coefficient was calculated.

***Figure 1 Rank correlation between TEC ranks and alternative weighting systems***




[15] It is also possible that the international reputation of VUW was based on the earlier 2006 PBRF, given the lower rank of VUW at that time.

It was found that ranks derived from the experiments were perfectly correlated with the ranks derived using the TEC weights, over a large range of dispersed weights and a smaller range of compressed weights. The lack of sensitivity is illustrated in Figure 1, which plots the Spearman rank correlation coefficient against the weight assigned to A, with other weights adjusted as described above. Only integer values of weights are considered. The range of weights assigned to QEC grade A between 9 and 16 gives exactly the same ranking as the TEC method.

In view of the considerable reputational value attached by universities to their rank position, any alteration in the rank order is regarded as very important. It is therefore perhaps not appropriate in this case to apply conventional significance levels to tests. Nevertheless, the horizontal dashed line in the figure gives the critical value for 2.5 per cent level of significance, of 0.886, for 6 degrees of freedom.[16]

### 4.3 Preliminary assessment of the source of improvements in *AQS*s

This subsection considers two potentially important sources of change. The first is the contribution of changes in the number of non-research-active staff (R-quality researchers) relative to research active staff in contributing to changes in the AQSs for each university between 2003 and 2012.

Figure 2 shows the estimated AQSs for each of the 8 New Zealand universities for the three PBRF rounds, including the individual researchers with R-quality scores. It shows the persistence in the rankings of the top and bottom groups of universities, commented on in Section 4.2. Although the increment in the AQSs for Auckland University of Technology (+2.30) and Massey (+2.06) are two of the three largest, they remain amongst the three universities with the lowest AQSs (with Lincoln). Four other universities (Auckland, Canterbury, Otago and Waikato) remain broadly similar in ranking. The exception is Victoria University of Wellington. The increment in the AQS for Victoria University is substantially higher than for others in the top group, and is the second highest (+2.30) among all the universities. The major proportion of this improvement to Victoria University's AQS took place between 2006 and 2012.

---

[16] The critical value for a significance level of 5 per cent is 0.829. Hence at this level, the correlation coefficients for all of the weights in Figure 1 exceed the critical value.

Figure 3 shows the estimated AQS for each university for the three rounds, when the individual researchers with an R-quality QEC are excluded. Figures 2 and 3 show that there has been an improvement in AQSs for all New Zealand universities between 2003 and 2012. They also reveal that the increases in AQSs when the R-quality researchers are excluded (Figure 3) are smaller than the increases in AQSs when the R-quality researchers are included (Figure 2). This suggests that the reduction in the number of submissions of non-research-active staff has been a significant factor contributing to the improvement in average quality of PBRF assessed researchers at New Zealand universities.

The changes observed in the AQSs for each university are the result of a substantial reduction in R-quality people, with a smaller contribution arising from improvements in the average quality of research-active staff (with letter scores of A, B, or C). But these aggregate changes are the result of several types of transition process. The reduction in the number of R-quality people could be the result of staff in these categories exiting the New Zealand university system, matched by fewer non-active researchers entering New Zealand universities. It may also be the result of non-active researchers transitioning to an active researcher status. Similarly, the changes observed in the AQSs for research active staff may be the result of several possible transitions, including changes in the quality of staff entering the system, staff transitioning to higher quality researcher status over time, and so on. These transitions are examined in Section 7.

*Figure 2 Average Quality Scores including Rs: 2003 to 2012*

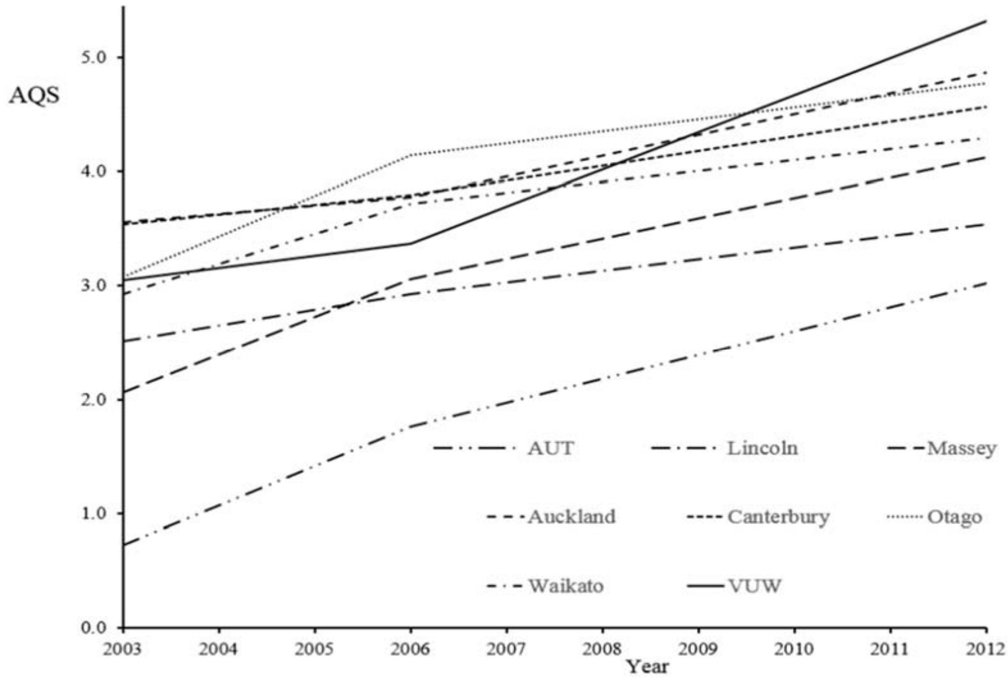
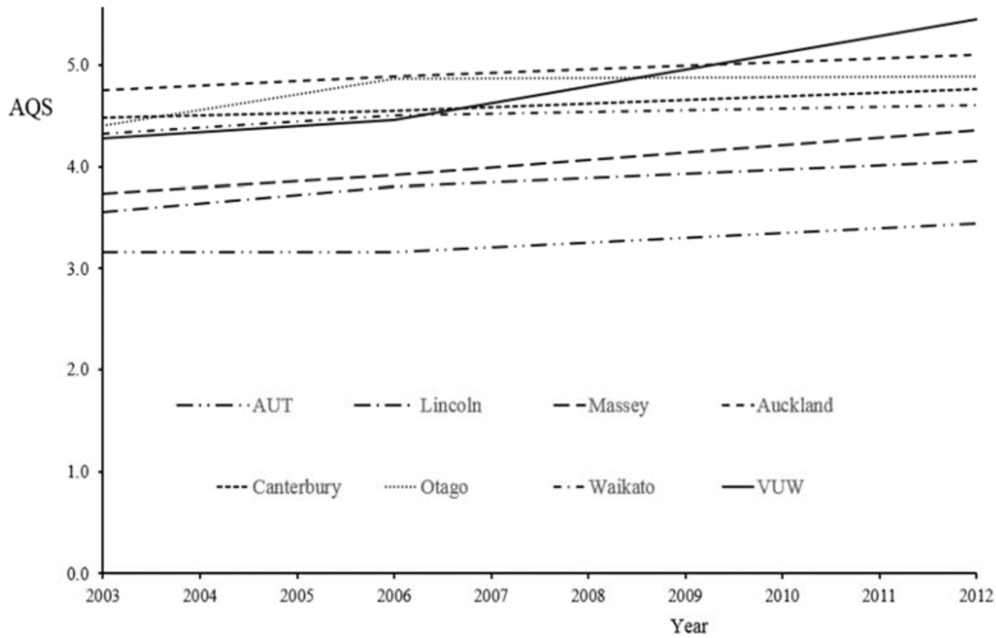



*Figure 3 Average Quality Scores excluding Rs: 2003 to 2012*



### 4.4 Alternative measures of average quality score

The previous subsection has shown the extent to which the AQS for a university depends on whether or not Rs are included in the denominator. In view of the complexity around the decision to submit a research portfolio, this subsection extends the analysis to consider all staff. This figure is not available in the current data, but it is possible to use the information provided in Table 1 to make an appropriate adjustment. This can be achieved as follows.

Let $Q$ denote the *total* quality score for a university. This is independent of whether R-quality researchers are included, since the score for Rs is zero. As above, $N_P$ denotes the number of people who submit a portfolio; $N_R$ denotes the number of R-graded portfolios and $N_{NP}$ denotes the number for whom there is no portfolio, with $N_T = N_P + N_{NP}$ representing the total number of non-administrative staff employed by the university. Ignoring weighting by fraction of a full-time employee, let $AQS_{(+R)}$ denote the average quality score including Rs, so that:

$$AQS_{(+R)} = \frac{Q}{N_P} \tag{3}$$

Let $AQS_{(-R)}$ denote the average quality score excluding Rs. Hence:

$$AQS_{(-R)} = \frac{Q}{N_P - N_R} \tag{4}$$

As discussed above, TEC reported values of $AQS_{(+R)}$ for 2003 and 2006, and $AQS_{(-R)}$ for 2012. Let $AQS_{(N_T)}$ denote the average quality score including all non-administrative staff. Hence:

$$AQS_{(N_T)} = \frac{Q}{N_T} = \frac{Q}{N_P + N_{NP}} \tag{5}$$

This can be computed by adjusting the other averages, since, for example:[17]

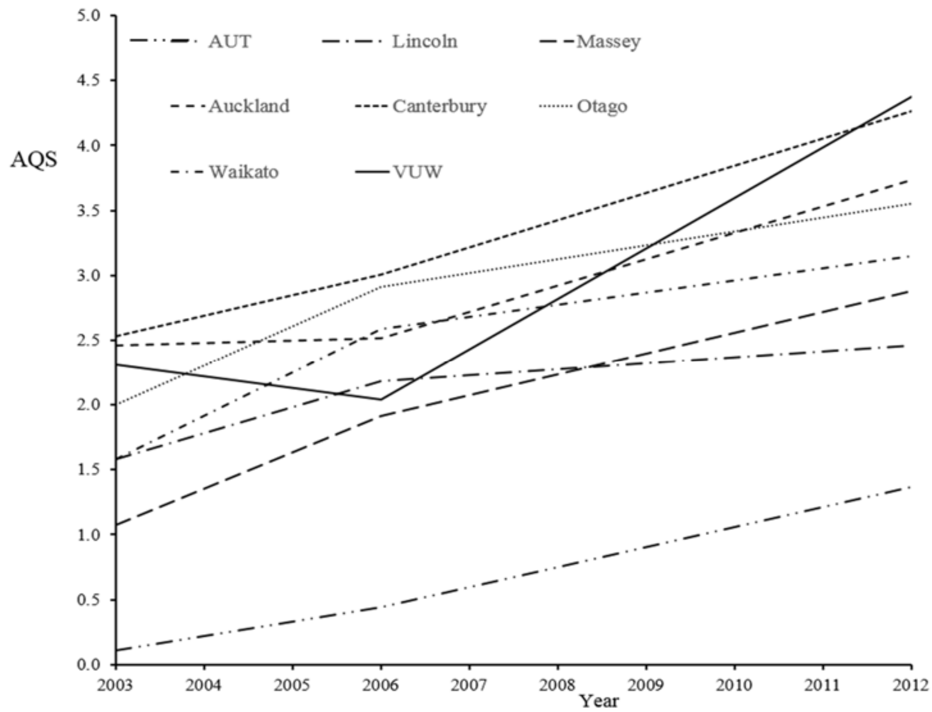
[17] Similarly: $AQS_{(N_T)} = AQS_{(-R)}\left(N_P - N_R\right) / N_T$.

$$AQS_{(N_T)} = AQS_{(+R)}\left(\frac{N_P}{N_T}\right) \tag{6}$$

Hence, using the values of $N_P / N_T$ from Table 1 (bearing in mind that these are obtained using weighted totals), Figure 4 shows the change in the AQS when all non-administrative staff are included. The absolute values of quality scores are lower than for Figures 2 and 3, but the general increase, with the exception of VUW in 2006, reflects the increase in the ratio, $N_P / N_T$ .

The use of all non-administrative staff as denominator for the AQS obviously changes the levels and also changes the ranking of the top four universities. Although VUW remains the top-ranked university in 2012, on this basis Canterbury has the second-highest AQS.[18]

***Figure 4 Changes in Average Quality Scores based on all staff: 2003 to 2012***




[18] The 2012 round took place after the 2011 severe Canterbury earthquakes. TEC agreed on a number of actions to mitigate the effect on academic staff at Canterbury and Lincoln universities: see New Zealand Tertiary Education Commission (2013a).

## 4.5 Assessing convergence

Table 3 and Figures 2 to 4 suggest there is some persistence in the ranking of the average research quality of researchers at New Zealand universities. Four universities were consistently positioned in the top four AQSs in each of the three rounds: Auckland, Canterbury, Otago and VUW. Waikato was ranked 5$^{th}$ in each round. AUT, Lincoln and Massey were ranked in the bottom three positions in all three rounds, with AUT being 8$^{th}$ in each round. However, it is worth considering in more detail whether there has been some convergence over the PBRF period in the actual AQSs, as opposed to the ranks.

A widely-entertained productivity and growth convergence hypothesis holds that in comparisons amongst countries, productivity growth rates tend to be inversely related to the initial level of productivity.[19] New Zealand universities face similar opportunities to acquire and apply management practices, access researchers from the global market, and are subject to the same Government tertiary policy and funding environment. If the productivity convergence hypothesis holds for relative rates of improvement of research capability in New Zealand universities, an inverse relationship between the proportional improvement of the sample of AQSs and initial AQSs would be expected.

Figure 5 plots the relationship between the proportional improvements over the period 2003 to 2012 in AQSs for each university against the respective initial (2003) AQS, where the latter are obtained by including all portfolios with a QEC rating of R. This diagram suggests there may be a tendency for university AQSs to converge. However, AUT is clearly exceptional in its proportional rate of increase, though of course it began with a very low AQS. Similarly, Massey's proportional rate of increase is higher than others, and also started with a slightly lower initial AQS than all but AUT. But aside from the observations for AUT and to a lesser degree Massey, there appears to be little tendency for the other universities to converge.

[19] See, for example, Baumol (1986) and Abramovitz (1986), where this inverse relationship is attributable to diminishing returns to the acquisition and application of knowledge.

*Figure 5 Proportional changes in AQS (including Rs): 2003 to 2012*

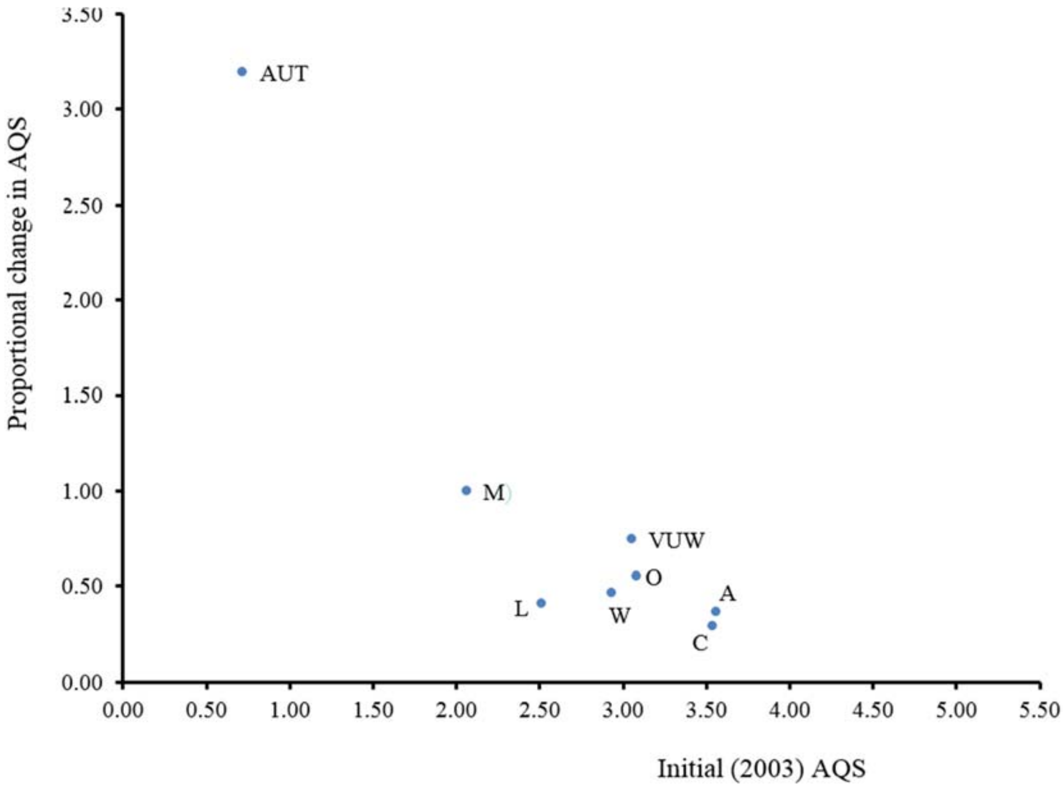
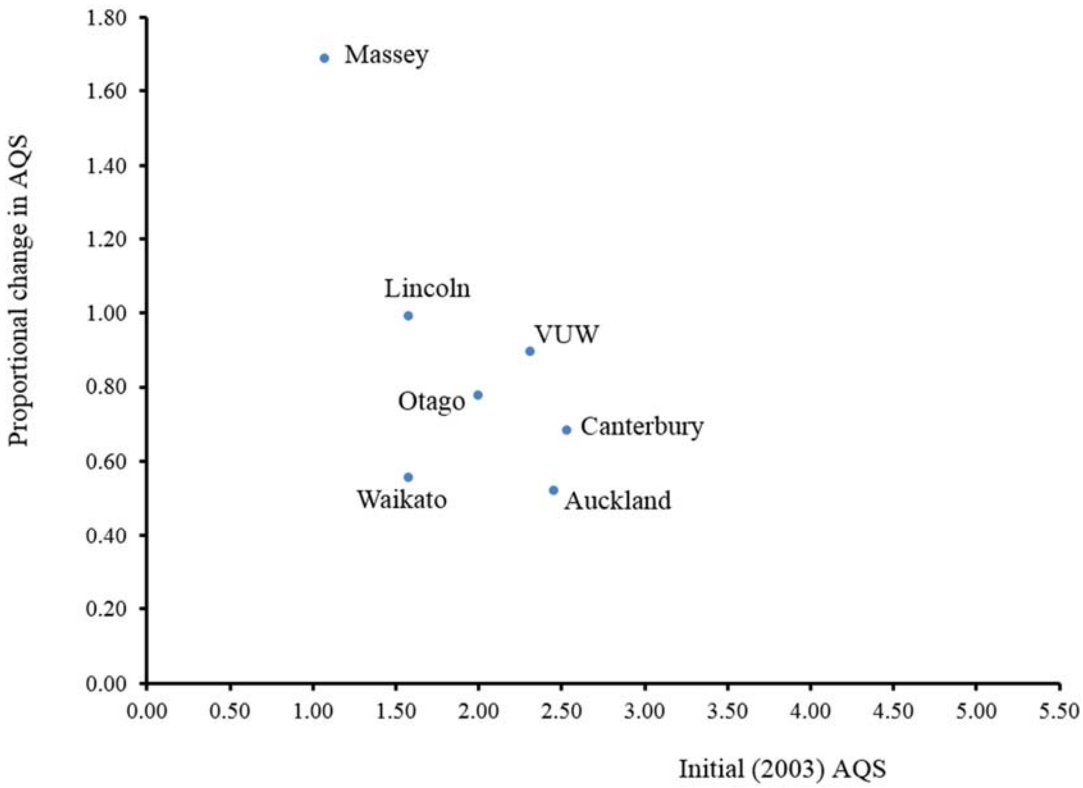



*Figure 6 Proportional change in AQS (including all staff)*



When an adjustment is made using all non-administrative staff in the denominator, as discussed in Section 4.4, the improvements in the AQSs are larger because of the increase in the proportion of staff who submit portfolios over the period. Figure 6 shows the proportional changes using these adjusted AQSs, although AUT is excluded because of its substantial increase of about 1100 per cent. With the additional exception of Massey, with an increase of about 170 per cent, the increases are in the range of about 50 to 100 per cent. These are substantial improvements for institutions in which structural change is generally regarded as being difficult. Further implications of such large changes are examined in Section 7 below.

***Table 4 Regression results for alternative convergence models***

| Coefficient | All | Exclude AUT | Exclude AUT and Massey | Top 5 by AQS |
|---|---|---|---|---|
| Intercept | 12.521 | 4.513 | -1.276 | 1.146 |
| (standard error) | (0.737) | (1.473) | (4.167) | (3.446) |
| (P-value) | (1.29E-05) | (0.038) | (0.779) | (0.771) |
| 2003 AQS | -12.636 | -3.612 | 2.197 | 0.049 |
| (standard error) | (1.122) | (1.685) | (4.241) | (3.441) |
| (P-value) | (9.63E-05) | (0.099) | (0.640) | (0.999) |
| 2003 AQS$^2$ | 3.212 | 0.837 | -0.571 | -0.099 |
| (standard error) | (0.392) | (0.454) | (1.043) | (0.836) |
| (P-value) | (0.00044) | (0.139) | (0.622) | (0.917) |
| $\bar{R}^2$ | 0.969 | 0.606 | -0.398 | 0.166 |
| F statistic | 111.756 | 5.614 | 0.289 | 1.398 |
| (P-value) | (7.08E-05) | (0.069) | (0.768) | (0.417) |

Note: These models are estimated using AQSs based on all staff as in Figure 4.

Convergence in AQSs can be examined more formally by estimating a quadratic relationship between the proportional change in AQS and the initial (that is 2003) AQS, and the square of the initial AQS. Results for alternative specifications are reported in Table 4. The quadratic convergence model that includes observations for all eight New Zealand universities generates results that are consistent with convergence to the mean AQS for all universities. However, it is evident from Figures 2, 4, 5 and 6 that AUT, followed by Massey, achieved the largest proportional improvements in their AQS between 2003 and 2012. When the quadratic convergence model is estimated with the observations for AUT excluded, the explanatory power of the model is significantly reduced. The coefficient on initial AQS is only significant at the 90 per cent level and the non-linear term is not significant at normally accepted levels. The model breaks down and no coefficients have the correct sign for convergence when it is estimated with the observations for both AUT and Massey excluded. When the model is estimated using only the observations for the top five ranked universities in 2012, none of the coefficients is significant at normally accepted levels of significance.

These tests suggest that the only source of PBRF convergence among New Zealand universities has come from improvements to the AQS of AUT and, to a lesser extent, Massey. Comparison of Figure 3 with Figure 2 suggests the improvements for AUT and Massey have arisen because they have achieved large reductions in the proportion of R-category staff.

## 5. Changes in the age distribution

This section examines changes in the age distribution of researchers and grades, and considers whether there are changes in the grade distributions by age. The sample used in this subsection to evaluate changes in age distributions is smaller than the sample used in the other subsections because the age of some researchers was not reported in the TEC dataset. The research capability of individuals is likely to be associated with their experience and therefore age. Accordingly, changes in the AQSs for each university are likely to be associated with changes in the age distribution of researchers assessed in each PBRF round.

Section 3 reported a fall of 5.5 per cent (or 388) in the number of PBRF researchers between 2003 and 2012. Three questions are considered in this sub-section. The first question considered is whether the reduction in the number of researchers between 2003 and 2012 is associated with a change in the age distribution of researchers. The second two questions considered are whether any observed change in the age distribution has resulted in a change in the age distribution by grade, and whether there has been a change in the grade distribution by age. The core data used to evaluate these age distribution issues is summarised in Table 5 which gives the joint age and grade frequency distributions over all universities for 2003 and 2012, for those individuals where a date of birth is recorded.

Table 6 summarises some of the features of Table 5, and shows that the fall in the number of PBRF researchers (with their age reported) declined by 6.4 per cent (441 researchers).

Table 6 also suggests 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 cohort, except the oldest cohort (60-89 age group). The largest absolute decline was in the 40-49 age cohort, 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 (153 researchers). The only age cohort to increase was the oldest cohort. The number of researchers in the 60-89 age cohort increased by 602 researchers, a 77.7 per cent growth rate for that age cohort.

***Table 5 Age and Grade Distributions in 2003 and 2012***

| Grade | Age Group | | | | | |
|---|---|---|---|---|---|---|
| | 20-29 | 30-39 | 40-49 | 50-59 | 60-89 | Total |
| *2003* | | | | | | |
| A | 0 | 32 | 119 | 200 | 91 | 442 |
| B | 4 | 328 | 641 | 585 | 200 | 1758 |
| C | 84 | 647 | 766 | 594 | 235 | 2326 |
| R | 150 | 507 | 754 | 744 | 249 | 2404 |
| Total | 238 | 1514 | 2280 | 2123 | 775 | 6930 |
| *2012* | | | | | | |
| A | 0 | 48 | 229 | 302 | 265 | 844 |
| B | 9 | 392 | 796 | 768 | 576 | 2541 |
| C | 71 | 775 | 757 | 706 | 469 | 2778 |
| R | 5 | 55 | 83 | 116 | 67 | 326 |
| Total | 85 | 1270 | 1865 | 1892 | 1377 | 6489 |

***Table 6 Change in the age distribution of PBRF researchers between 2003 and 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 a per cent of 2003 researchers in age group | -64.3 | -16.1 | -18.2 | -10.9 | +77.7 | -6.4 |

The development of research capability requires considerable investment and ‘time to build’. It is likely that the PBRF scheme has influenced university recruitment and retention policies in ways that have changed the age distribution. Accordingly, the changes observed in Tables 5 and 6 may be associated with changes in the age distribution by grade, and also the grade distribution by age. To evaluate whether such changes have occurred, the observed proportions for 2003 (obtained from Table 5) are used to derive expected frequencies for the 2012 age distribution by grade and 2012 grade distribution by age. A chi-squared goodness-of-fit test is then applied to compare the observed and expected frequencies for 2012 and test whether the 2012 distributions are significantly different from those of 2003.

The differences between the observed and expected frequencies for the 2012 age distribution by grade are shown in Table 7. The estimated chi-squared statistic is 608.85, which is well in excess the 1 per cent critical value of 26.2 for 12 degrees of freedom.

***Table 7 Differences between observed and expected 2012 age distribution by QEC***

| QEC | 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 |

The sum of each column in Table 7 shows the differences for each age cohort between observed and expected frequencies, and the sum of each row is zero. These totals reveal that the large increase in the frequency of researchers in the oldest age cohort (60-89) observed in Table 6 is significantly higher (by 599.71) than could be expected from the 2003 marginal frequencies. Conversely, the declines in the frequency of researchers in all the other (younger) age cohorts are larger than could be expected from the 2003 marginal frequencies. The largest unexpected changes are in the 'excess' decline of approximately 306 in the 40-49 age cohort, and the 'excess' rise of approximately 600 older age cohort researchers.

In the following discussion, the term 'unexpected change' refers to a difference between the actual 2012 frequency in a cell and the frequency expected if the form of the relevant distribution is the same as in 2003 and only the absolute (marginal) total has changed.

The higher than expected increase in older age cohort researchers is reflected in larger than expected increases in 60-89 age cohort people in all categories. The largest unexpected increases in this age cohort are in the research-active categories (A, B, and C), but there were also higher than expected numbers of R-category people. Conversely, there are large unexpected declines in the number of people in the lower QECs (C and R) in the 20-29 age cohort, of Bs in the 30-39 age cohort, of C and B researchers in the 40-49 age cohort, and very large unexpected declines in A and B researchers in the 50-59 age cohort.

The reasons for these unexpected changes are not clear. The unexpected increase in older people in the A and B-category may reflect changes in university retention and recruitment policies. They may also reflect changes in the support given to staff to enhance their research productivity thereby increasing the numbers transitioning to higher research categories. The unexpectedly 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. 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. The 'excess' decline in A, B, and C researchers in the middle-age cohorts may reflect their greater international mobility or reflect the effect of some transitioning to higher quality research QECs (including to A-grade). These issues are explored in more detail in Sections 6 and 7 below, where potential differences between universities are also explored.

***Table 8 Differences between observed and expected 2012 QEC distribution by age***

| QEC | Age | | | | | |
|---|---|---|---|---|---|---|
| | 20-29 | 30-39 | 40-49 | 50-59 | 60-89 | Totals |
| A | 0.00 | 21.16 | 131.66 | 123.76 | 103.31 | 379.89 |
| B | 7.57 | 116.86 | 271.67 | 246.65 | 220.65 | 863.40 |
| C | 41.00 | 232.27 | 130.43 | 176.63 | 51.46 | 631.79 |
| R | -48.57 | -370.29 | -533.76 | -547.04 | -375.42 | -1875.08 |

The differences between the observed and expected frequencies for the 2012 QEC distribution by age are shown in Table 8. The estimated chi-squared statistic is 2677.07, which is well in excess the 1 per cent critical value of 26.2 for 12 degrees of freedom.

The sum of each row in Table 8 shows the sum of the differences between observed and expected frequencies for each research category. The sum of the 'Totals' column is zero. There was a decline of 1875 larger than would be expected from a constant distribution in the number of Rs. This unexpected decline is evident in all age cohorts, but the smallest difference between the observed and expected frequencies is in the youngest age cohort; much larger differences occurred in all the other age cohorts. There were higher than expected frequencies in all the research-active categories (C, B, and A). There were large unexpected increases in C and B researchers in the 30-39 age cohort. The unexpected increases in the older age cohorts tend to display relatively higher increases in the higher research categories.

## 6. Hypotheses regarding organisational change

Faced with a need to improve the PBRF score of a department or faculty, several possible approaches exist, each with different costs and constraints. There are also institutional constraints in New Zealand, relating to hiring and firing of academics, the nature of contracts, the degree of wage flexibility, the allocation of university funds among departments, the concentration of power to make hiring and firing decisions, and so on. These features make the construction of a formal dynamic university model extremely difficult. Nevertheless, it is useful to consider the broad characteristics that may be expected to follow from an institutional objective of maximising, subject to constraints, the average quality score. In addition, it is necessary to consider how hypotheses might be tested in practice, given the limitations of the data.

A simple model is presented in subsection 6.1. Although this is highly simplified both in terms of objectives and constraints, it can help in thinking about possible responses to challenges facing universities. The next three subsections consider entrants, exits and transitions in turn. In each case the hypothesis is stated, followed by its rationale and brief discussion of how it could be tested.

It is trivially obvious that the maximum score is achieved when every person is graded as an A-type researcher. But this is also obviously neither feasible nor necessarily entirely desirable in practice, so a variety of strategies is likely to be observed. While, as mentioned above, the complexity of objectives and constraints militates against the construction of a formal organisational model and associated optimisation problem, thinking in terms of a general objective of maximising the score, subject to a range of financial and institutional constraints, suggests a number of useful hypotheses. These provide a guide or framework for the analysis of available data.

### 6.1 A Simple Model

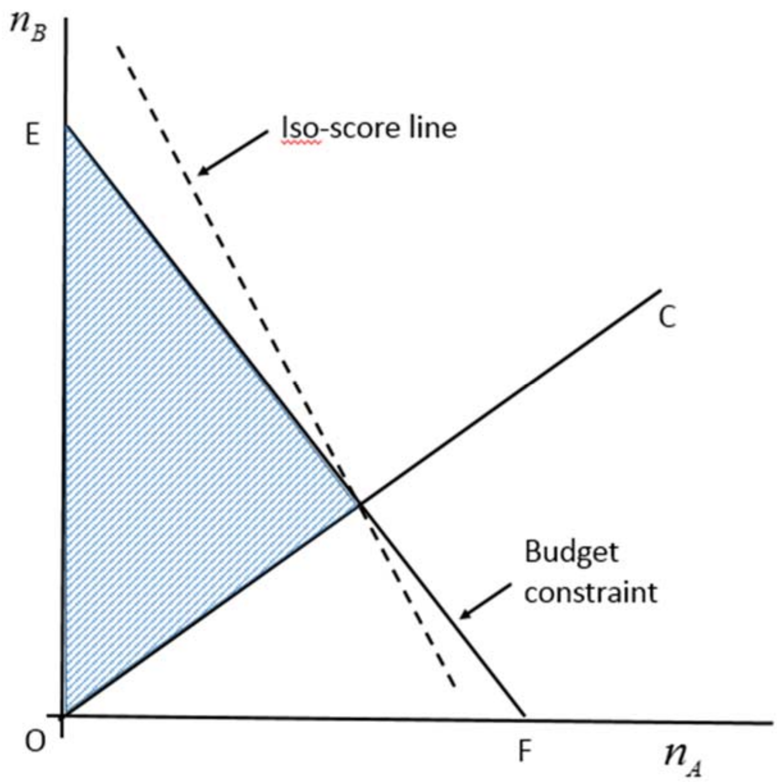
Consider a highly stylised model in which a university is considered to maximise its average quality score subject to a budget constraint and fixed 'prices'. Suppose there are just two types of researcher, A and B, and the number of each is equal to $n_A$ and $n_B$ respectively. The weights, or quality scores (equivalent to the numerical value of each QEC), attached to each are $s_A$ and $s_B$, so that the aggregate quality score is $W = s_A n_A + s_B n_B$. If the prices of each type are $w_A$ and $w_B$, the budget constraint is given by $R = w_A n_A + w_B n_B$. The university is assumed to maximise $W$ subject to the budget, $R$. The situation can be illustrated as in Figure 7.

The line EF represents the budget constraint and has a slope of $-\frac{w_A}{w_B}$. The dashed line in the figure represent an iso-score line, with a slope of $-\frac{s_A}{s_B}$. The aim is to move the iso-score line to the highest (most north-east) position, subject to the budget line. In this extremely simple linear model, the choice would involve a corner solution with only one type of researcher employed. Here the relative cost of As is low compared with the relative scores, so that the dashed iso-score line is steeper than the budget constraint: the outcome would be to employ only A-type researchers at the point F.

However, this neglects other potentially important constraints on the composition of researchers. Indeed, experience is required to attain the status of an A-type researcher, and there needs to be some 'entry grade' people, who can expect to transition to grade A. Indeed, this aspect was stressed by Tertiary Education Commission (2013a, p. 57), which suggested that, 'Just as a quality score between 8.00 and 10.00 is not realistically achievable (except by very small academic units); it is also not necessarily something to which it would be prudent to aspire. Indeed, any academic unit (or TEO) concerned about its longer-term viability and future research capability should have a strong interest in ensuring that it has within its ranks not only a sufficient number of experienced and well-respected researchers, but also a pool of new and emerging researchers'. Furthermore, there is a need to have sufficient staff to fulfil teaching commitments.

Hence, suppose there is an imposed ratio of As to Bs.[20] The feasible region is then the shaded part of the diagram and the optimal position is the corner solution where a mixture of As and Bs is employed. A reduction in the price of As leads to the employment of both more As and more Bs, with the mix depending on the slope of the line OC. Of course, these schedules are not likely to be straight lines, and the 'prices' of each type of researcher are not likely to remain constant as the demand increases. But more general shapes of the relationships lead to a diagram similar in form to that of Figure 7.

***Figure 7 A Simple Model of Quality Composition***




Suppose instead that the price of type-A researchers is high relative to the score attributed to them. This means that the dashed line in the figure is relatively flatter than the budget line, instead of being steeper. In this case only B-type researchers would be employed, at the corner E. It is simply not worth paying the relatively high price for any A-types. Of course, this is an extreme situation, but there are in practice a number of universities with

[20] Such quotients have in fact been an institutional feature of some universities in the past.

very few A researchers over a long period. The price of researchers is not necessarily constant across all universities. Some universities may, for a variety of reasons, find it very difficult to attract high-quality researchers. Universities with more graduate students and a high reputation may find it much easier to attract A-type researchers. This effect is also reinforced by the existence of a funding formula, as follows.

In practice, the budget constraint is affected by a funding formula which rewards higher quality. Suppose for simplicity that extra funds, $F(W)$, are available, depending on the value of $W$. Writing $\frac{dF}{dn} = f_n$ it can be seen that the slope of the budget constraint is no longer $-\frac{w_A}{w_B}$, but is equal to $-(w_A - f_{n_A}) / (w_B - f_{n_B})$. Hence on the reasonable assumption that $f_{n_A} > f_{n_B}$, the budget constraint becomes flatter as well as moving outwards. The funding formula therefore acts to reduce the effective price of A-type researchers, for those universities with higher AQSs. Hence it is possible that there could be a degree of polarisation, with two types of university. The higher quality researchers could become largely concentrated in a few universities.[21]

### 6.2 Entrants

**H1**: *More recruitment of As and Bs from outside are expected.*

As suggested in the previous subsection, while it is desirable to have a substantial number of A-grade researchers, this would result in a rather unbalanced age distribution, since As are more experienced researchers. Also, cost constraints are important. Nevertheless, the desire to improve the AQS relatively quickly implies that energy would be devoted to recruitment of A and also B people.

The data give ‘snapshots’ in the three relevant years, so information is not known about the precise date when each entry and exit to an organisation took place. However, to consider this hypothesis, entrants (or, rather, those recorded as providing a profile in an organisation for the first time) in 2006 and 2012 can be examined to see if the entrants by As and Bs are higher for higher-scoring universities and/or for those with larger increases in their AQS.

---

[21] The possibility that the PBRF could generate polarisation through the increased allocation of resources to higher-ranked universities was discussed by Curtis and Matthewman (2005).

**H2**: *An increase in the age of entry to the system, by Rs, is expected.*

A response to the risk that new researchers may not become sufficiently productive is to restrict recruitment to those who have already established a research and publication profile following the award of a PhD. The potential to rise to higher categories can then be judged more easily.

To examine this hypothesis, the age distribution of entries, particularly by Rs and R(NE)s can be considered. The distributions should reflect a shift towards entries by older R(NE)s and few entries by Rs.

### 6.2 Exits

**H3**: *Exits of Rs expected within higher age groups.*

An obvious method of increasing the PBRF score is to minimise the number of R members of each department. The rate of return from trying to convert an R into an A is clearly lower for older members. Hence, one mechanism is to encourage older R members to retire. However, given the nature of university contracts, this can present difficulties.

The data do not give reasons for exiting an organisation. A variety of reasons is possible. For example, exit could be through normal retirement, movement to another university (which may also be associated with achieving a higher QEC in the new university), failure to get tenure, or redundancy. However, to consider this hypothesis, the age distribution of those recorded as making exits can be compared for high-scoring universities and/or for those universities with relatively larger increases in their AQS.

**H4**: *Increased competition in the academic labour market by universities eager to recruit at higher levels may see more turnover at those levels*.

It was suggested above that more recruitment of high-quality researchers from other universities is likely. An obvious concomitant of this is that exits are also expected. This can be tested by examining the exits of Bs and As from lower-ranked universities and considering if these are higher than for the stronger universities.

### 6.3 Transitions

**H5**: *The upward movement of individuals is expected to increase for higher-performing universities as recruitment improves and as more of the older Rs are encouraged to leave.*

**H6:** *More upward movement is expected by R(NE) and C(NE) people compared with R and C people.*

A further method of increasing the number of As and Bs is to achieve internal growth. The improved selection of early-career individuals with high potential should itself see more movement up the scale, though this is less likely for older Rs. The use of, for example, internal research grants may promote more movement from C to B and B to A. Such movements are expected to be more frequent in higher-performing departments, although some outward movement of upwardly mobile individuals to other departments is also expected, as other universities compete for skills.

As mentioned above, information is not available about the transitions made by people between universities, so a number of transitions between quality categories are not recorded. However, it is possible to examine transition matrices to see if the higher-ranked universities are better at transforming Cs to Bs and Bs to As. Also, it is possible to consider whether R(NE) and C(NE) people are more likely to improve their QEC, compared with Rs and Cs.

### 6.4 Brief Summary of Evidence

Detailed evidence regarding the post PBRF experience is provided in the following section. However, a broad summary of the results regarding the various hypotheses discussed above, and motivated by the simple quality-maximising framework, can be given here. There is some support for H1, particularly among the five higher-ranking universities, but for the three lower-ranked universities the high 'effective cost' of recruiting at this level means that it is more beneficial to recruit more heavily at the lower levels.

Hypothesis H2 is supported by the evidence regarding age distributions, discussed in Section 5. It was also seen in Section 5 that H3 is rejected: relatively more Rs are in the higher age groups, despite the absolute reductions in their numbers. Furthermore, the median age in 2003 of those Rs who exited by 2012 is slightly less than the median age of other exits.[22]

[22] It is not possible to know the age at which transitions are made, as only the age at the time of a portfolio submission is known.

Regarding H4, it will be seen that there has in fact been relatively low turnover of A and B researchers, with the exception of Canterbury. But in this case, it is not clear to what extent outward mobility of the higher QEC people was related to PBRF relative to the earthquakes. Regarding H5, when it comes to converting researchers to higher-quality scores within the same organisation, this appears to have been very difficult, except for the higher-ranking universities. It is of course possible that the lower-ranked universities lose some of their more upwardly mobile researchers to other universities, and these are simply recorded as exits in the data. Hypothesis H6 receives support only with regard to R(NE) researchers.

## 7. Empirical results

The previous section has suggested a number of hypotheses regarding the possible responses to the incentive structure created by PBRF. In attempting to test these hypotheses, one immediate difficulty is that information is not available about research rankings or changes before the first PBRF.[23] However, it is possible to compare higher-performing universities with others, or universities with the largest improvements in their average score compared with those with smaller improvements. It may be expected that the various transitions vary by discipline. However, in view of space constraints, comparison among subject areas is a topic for future analysis.

### 7.1 All universities combined

It is first useful to consider the transitions among the various categories, along with entrants and exits, for all universities combined. These are shown for movements from 2003 to 2012 in Table 9. The flows are from rows to columns, and the transition proportions are given in parentheses immediately beneath the frequencies. The flows for those remaining in the same category (the diagonal entries in the matrix) are highlighted in bold.

[23] Several authors have endeavoured to assess the impact of PBRF on academic research productivity by comparing publication measures before and after the introduction of PBRF: 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) and, more generally, Hodder and Hodder (2010).

These overall flows 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.

***Table 9 Matrix of Flows: All Universities Combined 2003 to 2012***

| | Category in 2012 | | | | | |
|---|---|---|---|---|---|---|
| Cat in 2003 | A | B | C | R | Exits | Total |
| A | **242** | 71 | 2 | 0 | 127 | 442 |
| | (0.548) | (0.161) | (0.005) | | (0.287) | |
| B | 320 | **681** | 166 | 0 | 605 | 1772 |
| | (0.181) | (0.384) | (0.084) | | (0.341) | |
| C | 68 | 624 | **552** | 29 | 1101 | 2374 |
| | (0.029) | (0.263) | (0.233) | (0.012) | (0.464) | |
| R | 0 | 147 | 472 | **113** | 1721 | 2453 |
| | | (0.060) | (0.192) | (0.046) | (0.702) | |
| Entrants | 238 | 1099 | 1641 | 187 | | 3165 |
| | (0.075) | (0.347) | (0.518) | (0.059) | | |
| Total | 868 | 2622 | 2833 | 329 | 3554 | 10206 |
| | (0.085) | (0.257) | (0.278) | (0.032) | (0.348) | |

The largest exit rate, of just over 70 per cent, (referring to those classed as R in 2003 but who did not submit a portfolio for the same university in 2012) is for those who were classed as Rs in 2003. However, as discussed above, this proportion is somewhat misleading because some of the individuals could simply have been given new contracts which meant that they avoided the need to submit a portfolio. Nevertheless, low recruitment and high exit rates of Rs is a strong feature of transitions over the PBRF period.

Not surprisingly, a low proportion of entrants between 2003 and 2012 were classified as A-researchers in 2012. The majority of entrants are classed as Bs and Cs in 2012 (at 35 and 52 per cent respectively). Just under 20 per cent of the 2003 R's moved upwards to become Cs in the same institution by 2012, and 6 per cent moved upward to B, though again this may overstate the actual rate of improvement (since the denominator excludes those who did not submit a portfolio, although this was less important in 2003). Upward 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.

It also seems that of those classed as C-researchers in 2003, a high proportion, 46 per cent, had exited by 2012. It is perhaps 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 Bs also experienced high outward mobility, and a substantial proportion (29 per cent) of those who were As in 2003 had exited. This is consistent with a small proportion (7.5 per cent) 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.

These results suggest that the conversion of Bs and Cs to higher grades is less costly than that of Rs. This is consistent with the basic argument above that an improvement in the average quality score of a university is achieved by recruiting at the higher levels and appointing very few researchers with poor research outputs. As also seen above, this implies an increase in the average age of academics. The higher exit rate of Rs, along with a certain amount of mobility of A researches, as universities compete for higher-scoring academics, are both consistent with the discussion of the simple framework in Section 6.

It has been mentioned that additional QEC categories were introduced in 2006, to deal with relatively new researchers. A distinction was made between 'new and emerging' (NE) C and R staff and others. Table 10 shows the flows over this period for all universities combined: again, the diagonal entries are highlighted in bold. Of those who were graded as Rs in 2006 and submitted a portfolio in 2012, 7.6 per cent remained in the same university as R, 22.5 per cent moved upwards to C and 4.1 per cent moved to B (in the same university). However, of those who were R(NE) in 2006, 1.4 per cent became R, 7.4 and 28 per cent respectively were graded C(NE) and C in the same university in 2012, and 7.4 per cent were graded as B in 2012.

Of those who were graded C in 2006, 32, 27 and 1.5 per cent respectively stayed as C, moved to B, and moved to A in the same university by 2012. For those C(NE) researchers in 2006, 22.7, 27 and 1.8 percent respectively moved to C or C(NE), B and A by 2012.

***Table 10 Transitions 2006 to 2012 for all universities combined***

| Cat in 2006 | Category in 2012 A | B | C | C(NE) | R | R(NE) | Exits | Total |
|---|---|---|---|---|---|---|---|---|
| A | **355** | 125 | 2 | - | 0 | - | 142 | 624 |
| B | 317 | **966** | 238 | - | 0 | - | 592 | 2113 |
| C | 29 | 511 | **605** | - | 2 | - | 737 | 1882 |
| C(NE) | 14 | 208 | 169 | **5** | 42 | - | 330 | 768 |
| R | 0 | 41 | 223 | - | **75** | - | 651 | 990 |
| R(NE) | 0 | 36 | 140 | 36 | 7 | **0** | 269 | 488 |
| Entrants | 160 | 735 | 540 | 900 | 83 | 88 | - | 2506 |
| Total | 875 | 2622 | 1917 | 941 | 207 | 88 | 2721 | 9371 |

The distinction in the case of Cs appears therefore to have little value. But in the case of R researchers, those who were NE in 2006 do appear in general to have experienced more upward movement than other Rs. This supports the suggestion that the PBRF encouraged more careful selection of entry-level researchers by appointing people who had already established a minimum number of publications.

### 7.2 Individual universities

This subsection examines the extent of heterogeneity among universities in their recruitment and transitions. For example, it is of interest to consider how the experience of the universities with larger improvements in their AQS over time compare with those showing relatively little improvement. No definitive statement can be made in view of the absence of a non-PBRF counterfactual. However, it is possible to assess whether the experiences of those universities reflecting the most improvement are consistent with a ‘sharper’ incentive structure facing universities.

Table 11 provides information about the flows, from 2003 to 2012, between the different categories, along with the entrants and exits, for each university: the actual flows are shown in the left-hand side of the table. The transition proportions clearly differ among universities: these are easily calculated from the flows in the table, but are not reported here to save space. However, it is possible to test whether the pattern of transitions and exits differ (statistically) significantly from those obtained by taking all universities combined, as in Table 9. The right-hand side of Table 11 shows, for each university, the flows over the period which would result from starting with their actual stocks in 2003, and applying the transition and exit rates for all universities, as shown in Table 9. Using these hypothetical flows as 'expected' values, a standard chi-square test (based on 16 degrees of freedom) can be carried out.

The resulting values are reported in each case on the same line as the university name in Table 11. 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. Hence the only universities whose transitions do not differ significantly from those of all universities combined are Massey and Auckland. The university with the largest chi-square value is VUW, with 92.05: this had the highest AQS in 2012. The next-highest chi-square value is for AUT with 73.75: this university, as shown above, 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 322.75 is obtained.

This is the highest value in such pairwise comparisons: for example, comparing VUW with Auckland and Lincoln give chi-square values of 248.76 and 247.36 respectively. The lower-ranked universities also differ significantly from each other: hence comparing Lincoln with AUT and Massey give chi-square values of 53.48 and 170.07 respectively. There is clearly considerable heterogeneity among the New Zealand universities in their post PBRF experience.

Further inspection of Table 11 reveals that the universities also differ from each other in different ways.[24] Thus, AUT was relatively poor at recruiting As and Bs, and at converting 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 among the three universities that were consistently ranked at the low end of the AQSs, their experience differed. The common features are the low ability to recruit As and Bs, and

[24] 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 3..

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. In terms of the framework discussed earlier, it could be argued that these universities face relatively very high costs of recruiting As and Bs. Nevertheless, these universities managed to retain relatively more of their higher-scoring researchers.

Auckland was relatively 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 losses among its As and Bs. It was also relatively good 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 being relatively poorer regarding exits of Rs.

At the top of the AQS ranking in 2012, VUW achieved very good success at appointing As and Bs, keeping its higher-scoring staff, and it appointed relatively fewer Cs and Rs. Unusually, it had success in converting Rs to Bs. Among these top five universities, Canterbury stands out as the exception in terms of ability to appoint from people who were rated B researchers in 2012. Canterbury's higher losses of As and Bs may of course be explained to some extent by the earthquake experience, since those researchers are expected to be 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 As and Bs, and perhaps also more resources devoted to internal research grants, which stimulate higher productivity. Also, in recruiting Rs, they may also be selecting from those who are at the higher end of the scale. In terms of the framework discussed above, the greater ease of attracting As and Bs to the higher AQS universities, combined with the effect of the funding formula, involves an effective reduction in their price.

***Table 11 Flows from rows to columns for each university 2003 to 2012, along with those obtained if each university were to have the same transition proportions as all universities combined***

| | Actual | | | | | Hypothetical | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | A | B | C | R | Exits | A | B | C | R | Exits |
| **AUT** | | *Chi-square* =73.75 | | | | | | | | |
| A | 3 | 2 | | | | 3 | 1 | 0 | 0 | 1 |
| B | 3 | 18 | 1 | | 9 | 5 | 12 | 3 | 0 | 11 |
| C | 2 | 29 | 24 | 2 | 48 | 3 | 28 | 24 | 1 | 49 |
| R | | 11 | 84 | 22 | 357 | 0 | 28 | 91 | 22 | 333 |
| Entrants | 9 | 52 | 171 | 33 | | 20 | 92 | 137 | 16 | 0 |
| **Lincoln** | | *Chi-square* =50.54 | | | | | | | | |
| A | 5 | 1 | | | 1 | 4 | 1 | 0 | 0 | 2 |
| B | 7 | 17 | 6 | | 11 | 8 | 16 | 4 | 0 | 14 |
| C | 2 | 22 | 30 | 2 | 39 | 3 | 25 | 22 | 1 | 44 |
| R | | 0 | 10 | 9 | 41 | 0 | 4 | 12 | 3 | 42 |
| Entrants | 5 | 18 | 64 | 15 | | 8 | 35 | 53 | 6 | 0 |
| **Massey** | | *Chi-square* =17.67 | | | | | | | | |
| A | 25 | 6 | | | 13 | 24 | 7 | 0 | 0 | 13 |
| B | 34 | 83 | 27 | | 79 | 39 | 83 | 20 | 0 | 76 |
| C | 14 | 101 | 117 | 6 | 212 | 13 | 118 | 105 | 5 | 209 |
| R | | 34 | 93 | 23 | 430 | 0 | 35 | 112 | 27 | 407 |
| Entrants | 27 | 157 | 273 | 27 | | 36 | 168 | 251 | 29 | 0 |
| **Auckland** | | *Chi-square* =16.74 | | | | | | | | |
| A | 93 | 17 | 1 | | 48 | 87 | 26 | 1 | 0 | 46 |
| B | 115 | 192 | 45 | | 186 | 99 | 211 | 51 | 0 | 184 |
| C | 20 | 153 | 115 | 8 | 253 | 16 | 144 | 128 | 7 | 255 |
| R | | 29 | 86 | 20 | 279 | 0 | 25 | 80 | 19 | 290 |
| Entrants | 80 | 303 | 444 | 55 | | 66 | 306 | 457 | 52 | 0 |
| **Canterbury** | | *Chi-square* =59.06 | | | | | | | | |
| A | 24 | 9 | 1 | | 21 | 30 | 9 | 0 | 0 | 16 |
| B | 31 | 78 | 9 | | 95 | 34 | 73 | 18 | 0 | 73 |
| C | 5 | 59 | 49 | 0 | 140 | 7 | 67 | 59 | 3 | 117 |
| R | | 12 | 47 | 10 | 70 | 0 | 8 | 27 | 6 | 98 |
| Entrants | 20 | 124 | 173 | 17 | | 25 | 116 | 173 | 20 | 0 |
| **Otago** | | *Chi-square* =37.03 | | | | | | | | |
| A | 56 | 16 | | | 24 | 53 | 15 | 0 | 0 | 28 |
| B | 78 | 161 | 43 | | 101 | 76 | 161 | 39 | 0 | 131 |
| C | 14 | 127 | 130 | 1 | 207 | 14 | 126 | 111 | 6 | 222 |
| R | | 26 | 67 | 14 | 305 | 0 | 25 | 79 | 19 | 289 |
| Entrants | 43 | 211 | 313 | 14 | | 44 | 202 | 301 | 34 | 0 |
| **Waikato** | | *Chi-square* =57.33 | | | | | | | | |
| A | 12 | 13 | | | 9 | 19 | 5 | 0 | 0 | 10 |
| B | 18 | 59 | 25 | | 52 | 27 | 58 | 14 | 0 | 53 |
| C | 5 | 61 | 54 | 6 | 69 | 6 | 51 | 45 | 2 | 90 |
| R | | 9 | 51 | 13 | 109 | 0 | 11 | 35 | 8 | 128 |
| Entrants | 12 | 76 | 84 | 16 | | 14 | 65 | 97 | 11 | 0 |
| **VUW** | | *Chi-square* =92.05 | | | | | | | | |
| A | 24 | 7 | | | 11 | 23 | 7 | 0 | 0 | 12 |
| B | 34 | 73 | 10 | | 72 | 32 | 67 | 16 | 0 | 65 |
| C | 6 | 72 | 33 | 4 | 133 | 7 | 65 | 58 | 3 | 115 |
| R | | 26 | 34 | 2 | 130 | 0 | 12 | 37 | 9 | 135 |
| Entrants | 42 | 158 | 119 | 10 | | 25 | 114 | 171 | 19 | 0 |

## 8. Sustainability of Transitions

It has been stressed that these transitions relate only to the period during which PBRF has operated, and therefore cannot unambiguously demonstrate any effect of the research appraisal exercise itself. Nevertheless, there is no doubt that New Zealand universities in 2003 had a large number of Rs that was inconsistent with the incentive structure created by the PBRF exercise. Indeed, the high exit rate (of those categorised as R in 2003) and 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. That is, the question arises of whether this large improvement could be sustained over a long period.

This question may be considered more formally by considering the implied equilibrium distribution generated by the transitions and entries observed over the PBRF period. 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 just balanced by inward movements each period. In general, let *n* and *b* denote vectors of the number of individuals in each category and the number of entrants respectively, and *T* represents a square transition matrix. Then for a constant entry vector, *b*:

$$n_t = b + Tn_{t-1} = b + T\left(b + Tn_{t-2}\right) \qquad (7)$$

and so on. The power of *T* eventually goes to zero, so that the long-run equilibrium vector of numbers in each category is given, where *I* denotes the unit matrix, by:

$$n = \left(1 + T + T^2 + T^3 + \ldots\right)b = (I - T)^{-1}b \qquad (8)$$

Using the transition matrix and vector of births in Table 9 gives equilibrium stocks of A, B, C and R researchers respectively of 1519, 2708, 3138 and 999. This is clearly very different from the totals reported for 2012 in Table 9. It is therefore not unreasonable to 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.

***Table 12 Initial (2003) and Equilibrium QEC Distributions***

| | AUT | Lincoln | Massey | Auckland | Canterbury | Otago | Waikato | VUW |
|---|---|---|---|---|---|---|---|---|
| | *Long-run equilibrium number in each QEC* | | | | | | | |
| A | 143 | 42 | 186 | 388 | 115 | 350 | 140 | 234 |
| B | 121 | 50 | 391 | 709 | 241 | 616 | 206 | 349 |
| C | 119 | 47 | 498 | 835 | 287 | 668 | 214 | 262 |
| R | 204 | 85 | 135 | 292 | 145 | 167 | 93 | 105 |
| | *Number in each QEC in 2003* | | | | | | | |
| A | 5 | 7 | 44 | 159 | 55 | 96 | 34 | 42 |
| B | 22 | 31 | 157 | 400 | 139 | 306 | 111 | 128 |
| C | 105 | 95 | 450 | 549 | 253 | 479 | 195 | 248 |
| R | 474 | 60 | 580 | 414 | 139 | 412 | 182 | 192 |

Table 12 presents the long-run equilibrium QEC distributions for each university, along with the actual initial 2003 distributions. It is clear that the changes made since the introduction of PBRF in 2003, and up to 2012, are not sustainable in the long term. The extreme case is AUT, which has been seen to have by far the largest proportional increase in its AQS over the period. This arises because of its relatively high entry into the A and especially C categories, along with the high rate of transitions to Bs and the absence of any exits from the A category. Hence the only movement out of the A group consists of those moving to B. But, for all universities the equilibrium distributions are completely unrealistic.

These results help to provide a stark 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. They can also perhaps provide indirect support for the idea that the changes do represent, to some extent, the influence of PBRF.

## 9. Conclusions

This study used a TEC database consisting, for each university, of Quality Evaluation Categories (QECs) for each anonymous portfolio, with information about age and research discipline. For entrants and exits to and from each university, information about the origin and destination is not known. Similarly, the date of any movement between universities or into and out of the system as a whole is not known. Despite these limitations, the dataset provides a rich source of information with which to assess the nature of changes in the quality of researchers at New Zealand universities over the PBRF period.

The PBRF was introduced along with a change in the way research funding is allocated to universities, with the aim of improving incentives facing universities to improve their research performance. The decision by TEC not to request portfolios from all non-administrative staff, and leaving the number of those with QEC of R out of the calculation of AQSs in 2012, weakened the incentives facing universities. Nevertheless, there was an increase over the period 2003 to 2012 in the proportion of all academic staff submitting a portfolio, for all universities. Furthermore, a curious feature of the PBRF exercise is the use of QECs (given cardinal scores of 10, 6, and 2 for A, B and C respectively) rather than the raw scores (which range from 0 to 700) for each portfolio. The 'transformation' of the raw scores could influence the university rankings.

The way in which AQSs were calculated changed in 2012, so that published values distort the true change, but when values are adjusted to include all non-administrative staff, all universities had substantial proportional improvements in (adjusted) AQSs over the period. The only university experiencing a fall was VUW from 2003 to 2006, but this university made a very strong recovery to lead the rankings by 2012. However, there is little evidence of a process of convergence. The universities fall broadly into two groups consisting of the top 5 (which experienced rank changes) and the bottom 3 (again with within-group rank changes). There was little movement between these two groups. This may be affected by the self-reinforcing nature of the funding formula used by TEC: this reduces the effective price of A-researchers for those universities which are already better-placed to attract top researchers.

Initially, there was a large number of R-category faculty in the system. A dominant feature of the changes is the reduction in Rs in all universities. The initial number was particularly high in education, and a large proportion of exits came from this group, combined with a small number of entrants into education. 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 both the age distribution within grades and the grade (QEC) distribution within age groups. However, a relatively high rate of exit from older Rs was not observed.

There was a relatively high recruitment of As and Bs from among the five higher-ranking universities, with relatively low exit rates from the group (VUW being most successful regarding both these flows). However, the lower-ranked universities experienced relatively higher losses from the higher QEC academics. More recruitment was found among the lower QECs in the lower-ranked universities

Regarding transitions between QECs within the same university, the modal proportion was for ‘stayers’ in the same category, for all grades except for the movement from C to B. For those remaining in the same university, there were relatively few downward movements from As and Bs (in each case about 10 per cent of those starting in each grade in 2003).

A fundamental difficulty in evaluating the possible influence of the changed incentive structure created by PBRF is that comparable information is not available for a period before its introduction in 2003. Nevertheless, the proportional increase in adjusted quality scores ranged between about 50 per cent to about 100 per cent for most universities; exceptions were Massey, of about 170 per cent, and AUT, with over 1100 per cent (going from a very low AQS base). But even for higher-scoring universities large proportional changes were achieved, and this is particularly evident for VUW from 2006 to 2012.

In terms of absolute AQSs, it has been seen that the TEC scoring method distorts the measured overall increase in the quality of researchers over the period. This is because in 2012 the R-category researchers were excluded from the denominator in calculating the average. However, the fact that not all academic staff were included in any year means that the substantial increase in the proportion of total staff submitting a portfolio over the period was not reflected in the reported AQSs. Using a consistent denominator, and supposing that all those who did not submit a portfolio were in fact R-category academics, there were only four universities with adjusted quality scores above 2 in 2003. By 2012, despite the gains discussed here, no university had an adjusted average score above 4.5. This may be placed in perspective by recalling that an AQS of 2 (for C researchers) reflects ‘local standing’, while a score of 6 (for B researchers) reflects ‘national standing’. The maximum possible score, for ‘international standing’ is 10.

Although a number of research evaluation exercises have been carried out in other countries, the PBRF exercise is unique in its scoring method and its analysis of individual portfolios. Hence, it is unfortunately not possible to compare the New Zealand universities with those elsewhere. However, some perspective may be provided by reflecting on the number of universities in other English-speaking countries, compared with 8 in NZ. Australia and the UK have 43 and 147 respectively, while Canada and the US respectively have 98 and 2,618 universities.[25] It would be interesting to consider the proportion of these universities which have 'international standing'.

It is hard to imagine that the improvement over the 2003 to 2012 period would have arisen in the absence of a change to the incentive structure facing universities. Furthermore, if the changes (as reflected in the transition rates between QECs, and the entrants to each category) are held constant and projected forward, they give rise to completely unrealistic 'equilibrium' distributions. These findings provide support for the conclusion that PBRF did indeed contribute to raising the average research quality of NZ academics over the period. Nevertheless, it is impossible to be precise about the nature of its contribution. Furthermore, it is recognised that some of the improvement may well have arisen simply from a learning process about how to write persuasive evidence portfolios combined with investment by universities in providing advice and feedback.[26]

Nevertheless, it was suggested above that the introduction of PBRF probably did lead to significant increases in the average research quality of New Zealand university faculty. A substantial contribution to this increase was the large reduction in non-research-active staff and a greater emphasis on recruitment at higher levels. Although the changes over the period 2003 to 2012 do not appear to be sustainable over a much longer period, the question arises of what kind of incentive structure is appropriate following the changes that have taken place over the last decade. Consideration of the full implications and costs of the PBRF exercise, and possible alternatives, are outside the scope of the present study. In the absence of a fully-functioning academic labour and research market, some kind of central evaluation and reward (or pricing) system appears to be necessary: the form this should take presents a substantial challenge. The present analysis hopes to inform this wider debate.

[25] See https://www.studyinaustralia.gov.au/global/australian-education/universities-and-higher-education/list-of-australian-universities and https://en.wikipedia.org/wiki/List_of_universities_in_the_United_Kingdom and http://learningenglish.voanews.com/a/a-23-2005-05-11-voa1-83125492/124600.html and https://en.wikipedia.org/wiki/List_of_universities_in_Canada

[26] This point was also made by the Tertiary Education Commission (2013a, p. 66).

## References


Abramovitz, M. (1986) Catching up, forging ahead, and falling behind. *Journal of Economic History*, 46, pp. 385-406.

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.

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.

Bakker, L., Boston, J., Campbell, L., & Smyth, R. (2006). *Evaluating the Performance-Based Research Fund: Framing the Debate.* Wellington: Institute of Policy Studies.

Baumol, W.J. (1986) Productivity growth, convergence, and welfare: what the long-run data show. *American Economic Review*, 75, pp. 1072-1085.

Boston, J. (2007) The performance-based research fund: issues and options for the future. Victoria University of Wellington Institute of Policy Studies.

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*, 24, pp. 55-84.

Butler, P. and Mulgan, R. (2013) Can academic freedom survive performance based research funding. *Victoria University of Wellington Law Review,* 44, pp. 487-520.

Chavarro, D. Tang, P. and Ràfols, I. (2016) Why researchers publish in non-mainstream journals: training, knowledge bridging and gap fillings, *University of Sussex Social Policy Research Unit Working Paper Series, 2016-22*, December. Available at https://www.sussex.ac.uk/webteam/gateway/file.php?name=2016-22-swps-chavarro-et-al.pdf&site=25

Clear, T. and Clear, A. (2012) PBRF 2012 and the nature of degree teaching. Paper presented at 3rd Annual Conference of Computing and Information Technology Research and Education New Zealand (CITRENZ2012).

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

Curtis, B. (2004) The PBRF and its implications for academics (in the humanities and social sciences). Paper presented to Conference on *Labour, Employment and Work in New Zealand* held at Victoria University of Wellington. Available at: https://ojs.victoria.ac.nz/LEW/issue/view/144

Curtis, B. and Matthewman, S. (2005) The managed university: the PBRF, its impacts and staff attitudes. *New Zealand Journal of Employment Relations*, 30, pp. 1-17.

Duncan, G. (2008) Counting the currency of knowledge: New Zealand's performance-based research fund. In *Knowledge as Value: Illumination Through Critical Prisms* (edited by I. Morley and M. Crouch), pp. 23-42. New York: Rodopi.

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.

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, pp. 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.

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

KPMG (2012), *Performance-Based Research Fund Report on the overall preparedness of Tertiary Education Organisations for the Quality Evaluation*, Wellington, KPMG, pp 23. Accessed from http://www.tec.govt.nz/assets/Reports/PBRF-TEO-Preparedness-Report2.pdf

Mahoney, P. (2004) Performance-based research funding. Wellington: Parliamentary Library Background Note, 2004/07.

Ministry of Education (2008) How the PBRF has shifted research funding. Wellington: Ministry of Education.

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

Ministry of Education (2013a) Review of the Performance-Based Research Fund: Consultation Document. Wellington: Ministry of Education. Available at: https://education.govt.nz/assets/Documents/Further-education/Policies-and-strategies/Performance-based-research-fund/PBRFConsultationDocument.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).

New Zealand Tertiary Education Commission (2013a) *PBRF 2012 Final Report*. Wellington: Tertiary Education Commission.

New Zealand Tertiary Education Commission (2016) Performance-based research fund (PBRF) user manual. Wellington: Tertiary Education Commission.

OECD (2010) *Performance-Based Funding for Public Research in Tertiary Education Institutions: Workshop Proceedings*. OECD Publishing: http://www.oecd-ilibrary.org/education/performance-based-funding-for-public-research-in-tertiary-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. and Engler, R. (2013) *In Pursuit of Excellence: Analysing the Results of New Zealand's PBRF Quality Evaluations*. Wellington: Ministry of Education. Available from: www.educationcounts.govt.nz

Whitley, R. (2008) Changing governance of the public science. In *The Changing Governance of the Sciences: The Advent of Research Evaluation Systems* (edited by R. Whitley and J. Glaser), pp. 14-16.

Walsh, P. (2004) *Transcript of Plenary Sessions, PBRF Forum: Evaluation of the Assessment Framework*, The Royal Society of New Zealand Social Sciences Committee, http://www.rsnz.govt.nz/advisory/social_science/media/pbrf2004-plenary.php.

## 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

Chair in Public Finance
Victoria Business School

Working Papers in Public Finance