A Schumpeterian Gale: Using Longitudinal Data to Evaluate Responses to Performance-Based Research Funding Systems

Robert A. Buckle, John Creedy and Ashley Ball

WORKING PAPER 07/2020
December 2020

Working Papers in Public Finance

Victoria University of Wellington
Te Herenga Waka

Chair in Public Finance
Wellington School of Business and Government

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:
https://www.wgtn.ac.nz/sacl/centres-and-chairs/cpf/publications/working-papers

A Schumpeterian Gale: Using Longitudinal Data to Evaluate Responses to Performance-Based Research Funding Systems

Robert A. Buckle, John Creedy and Ashley Ball

Victoria University of Wellington

Abstract

Performance-based research funding systems (PBRFS) have been introduced in many countries for allocating funding to research institutions. There continues to be considerable debate about the effectiveness and consequences of these systems. This paper provides a new approach to this debate. It utilises longitudinal researcher data available from the New Zealand PBRFS, which assesses institutional performance and allocates funds based on individual researcher performance. The longitudinal data enable identification of entry, exit and quality transformation of researchers and the contribution of these dynamics to changes in university and discipline research quality, in a manner similar to Schumpeter's description of the impact of firm dynamics on productivity and economic growth, in terms of a 'gale of creative destruction'. The approach enables a deeper understanding of individual and institutional responses to PBRFSs, the sustainability of changes, and the contributions of changes in researcher quality and discipline composition to changes in institutional performance.

Key words: Performance-Based Research Funding Systems; policy evaluation; research quality; Social Accounting Framework; longitudinal data.

JEL classifications: I2; I28.

Acknowledgements: We are grateful to the New Zealand Tertiary Education Commission (TEC) for providing the data and thank Sharon Beattie, Morgan Healey, Paul Lawrence and Susan McFadyen for helpful discussions and support. We have benefited from comments on an earlier draft by Norman Gemmell. The research for this paper was supported by a Victoria University of Wellington Research Grant.

1. Introduction

Performance-based research funding systems (PBRFS) have been implemented in many countries during the last three decades as a basis for allocating funds to universities.[1] These schemes are designed to enhance research output, quality, and impact, and to strengthen accountability for the use of public funds. Three fundamental reasons influence the choice of system for public funding of research institutions: the need to ensure stable and reliable long-term financing; the need for accountability; and the need to incentivise performance. These motives, and especially accountability and incentive requirements, were associated with policy reforms in many countries which granted increased autonomy to publicly-funded universities, and spawned the implementation of performance-based research funding systems.[2]

PBRFSs vary by coverage and assessment methods which may be based on peer review, metrics and bibliometric data, or a combination of these (Guena & Martin, 2003; OECD, 2010; Wilsdon *et al.*, 2015). Despite their widespread use, de Boer *et al*. (2015, p. 5) suggest that, 'there still is not sufficient evidence on the effects of the systems and that our understanding of the proper design and implementation of performance agreements is still incomplete.'[3] Debate about the impact and merits of alternative designs continues.

A fundamental problem arises when trying to evaluate a PBRFS, because comparable data generally do not exist for the years before its implementation. This absence even makes it possible for some commentators to argue that external factors and existing incentives (applying to individual researchers and university managers) have long been in existence, and PBRFS have not actually led to changes that would not have taken place in the absence of the schemes. One approach, attempting to overcome this problem, is to examine a time-series of bibliometric data, used as a proxy for research output and quality, and to search for a structural change. However, it is argued here that an evaluation needs to pay attention to the

[1] These systems have typically been introduced as a component of a 'dual-support' system of funding university research. The performance-based funding component is allocated to a university by Government, based on results of a research performance assessment process, and typically the university has discretion over the use of those funds. Prior to the introduction of performance-based funding systems, such funding was often provided on a bulk funding process based on student numbers. The second component of the dual-support system is contestable research funding sought from independent research funding bodies and earmarked for specific research projects.

[2] An early and influential example is the UK's 'Research Excellence Framework' (REF) introduced in the 1980s. It was initially known as the 'Research Selectivity Exercise'. By 1989 the name was changed to 'Research Assessment Exercise', and since 2010 it has been known as the REF. Martin (2011) provides a brief history of the motivation for and early development of the REF.

[3] A similar conclusion was drawn by Checchi *et al.* (2019).

incentive structure created by the PBRFS, and to consider the precise nature of the dynamics of change. Such an evaluation is possible when longitudinal micro-datasets are available, recording the specific performance details of individuals over several rounds of a PBRFS.

The purpose of this paper is to demonstrate methods that can be applied to PBRFSs, and to improve understanding of the processes that change the perceived research quality of universities and the effects of the imposed incentive structures, using longitudinal micro-data. The methods, and in particular the adoption of a social accounting framework, are directly applicable to those schemes which involve the evaluation of individual researchers (and thereby use an explicit aggregation rule to obtain a measure for each unit), rather than simply considering university departments or other research units as a whole.

The change in the perceived research quality of a university arises from several types of transition. These include the entry, exit and quality transformations of researchers within the institution. They can also involve changes in the discipline composition of the university. In this respect, understanding how the research quality of a university or discipline group changes is in some ways analogous to Schumpeter's (1942) famous description of the way firm dynamics influence productivity and economic growth. Schumpeter emphasised that economic development involves an organic and evolutionary process, which amounts to 'reforming the economic structure from within, incessantly destroying the old one, incessantly creating a new one' (1942, p. 82). The present paper builds on this analogy. It presents techniques that can reveal some of the process of university reforms following introduction of a PBRFS by identifying the contributions to the growth in research quality of the entry, exit and quality transformation of researchers and changes in discipline composition. The methods are illustrated in the context of the New Zealand PBRFS, but can be applied to any system using individual assessment. They demonstrate the considerable advantages of access to longitudinal data, rather than being limited to several cross-sectional datasets recorded at the time of each evaluation round.

Section 2 provides a brief review of the recent international research evaluating the impact of PBRFSs. Section 3 explains the essential features of the NZ system, its key metrics, and the longitudinal researcher data used, covering all universities and collected as a result of the introduction in the early 2000s of a PBRFS.[4] The incentives created by the introduction of the

[4] An earlier series of paper (Buckle and Creedy, 2019a, 2019b, 2020, and Buckle *et al.*, 2020) examined the changes over the period 2003 to 2012. The present paper covers the complete period 2003 to 2018, and the dataset contains additional information about researchers' characteristics, including full-time equivalent status, and transfers among universities and disciplines.

system are discussed in Section 4, together with some hypotheses regarding organisational changes. Section 5 explains how the data can be utilised to trace changes in researcher quality and the entry, exit and quality transformation of researchers, and how these dynamics contribute to changes in overall university research quality. Section 6 applies a decomposition method to identify the separate contribution of these components to changes in the research quality of the entire university system and individual universities and disciplines. This is extended in Section 7 to identify the contributions of changes in discipline composition. The approach is adapted in Section 8 to assess the possible sustainability of observed organisational changes. Section 9 provides concluding comments.

2. Recent empirical assessments of the impact of PBRFSs

The introduction of the UK's performance-based funding system in 1986 prompted early contributions to a growing body of literature assessing responses to PBRFSs and whether their intended outcomes were achieved. The large number of research papers describing and evaluating the impact of the UK system illustrate the range of issues that have provoked international debate about the merits of these schemes; see, for example, Hare (2003), Barker (2007), Adams and Gurney (2010) and Broadbent (2010).

A critical question concerns the impact of PBRFSs on the volume and quality of research, although evaluation has typically been challenged by the absence of appropriate control groups. There are some quantitative cross-country comparison studies, such as Auranen and Nieminen (2010) and Franzoni *et al*. (2011). However, this type of approach poses several methodological difficulties, associated with identifying the precise timing of policy change, the diversity of schemes, and the difficulty of identifying the timing and potential influence on performance change of other factors.

An alternative popular approach has been to use a bibliometric index of research output to compare performance before and after the introduction of a PBRFS. Country studies using this method include Butler (2003), Adams and Gurney (2010), Anderson *et al*. (2013), Anderson and Tressler (2014) and van Besselaar, *et al*. (2017). Wang and Hicks (2013) take this approach a step further and use time-series techniques to search for a structural break in the time series of a bibliometric indicator of research output. They compare the estimated timing of a structural break with 'narratives of policy change', to determine whether the policy could have contributed to the observed change in the indicator. They apply this method to the UK's sequence of PBRFS rounds although, in principle, this approach could also be applied in a cross-country analysis. In a multi-country study, Checchi *et al.* (2019) also use 'narratives of policy change' to judge the timing of a policy-change dummy variable in a regression analysis of bibliometric-based indicators of research output. The significance of a policy dummy variable is estimated by a cross-sectional fixed-effects model, controlling for the type of assessment method (whether peer-review or bibliometric). In addition to the 20 countries in their sample, which introduced a national PBRFS, they include 11 countries that had not introduced this type of system during the sample period (1996-2016).[5]

Bibliometric indicators of research performance have also been applied using a production function approach to measuring university research performance. An early example of this approach is Moore *et al*. (2002), who assessed the productivity of academic researchers following the introduction of the UK system. Smart (2009) and Gemmell *et al*. (2017) use a variety of bibliometric indices of research output in their estimates of the productivity of New Zealand universities before and after the introduction of the PBRFS.

These bibliometric indicators have limitations which are recognised by many of the studies cited above.[6] A critical question is whether there is concordance between the bibliometric indices and the type of research output that national PBRFSs are designed to promote, measure and evaluate. This has been addressed by Ancaiani *et al*. (2015) in the context of the Italian research assessment exercise, called Valutazione della Qualità della Ricerca (VQR). The VQR considers research outputs, which include journal articles, books, book chapters and conference proceedings, patents, prototypes, project plans, software, databases, exhibitions, works of art, compositions and thematic papers. It uses a dual system in which both peer review and bibliometrics are considered. Although ANVUR, the Italian agency administering the VQR, concluded from an analysis of 10 per cent of the submitted research products that there is a 'more than adequate concordance' between evaluations carried out by peer review and those by bibliometrics (ANVUR, 2013 pp. 25-26), the method used to derive this conclusion has been contested by Baccini and De Nicolao (2016). Aksnes and Taxt (2004) undertook a similar type of investigation using research output at a Norwegian university and

[5] The papers by Wang and Hicks (2013) and Checchi *et al.* (2019) illustrate the type of bibliometric data used in this approach. Wang and Hicks measure research output by the number of Thomson Reuters WoS indexed journal publications; Checchi *et al.* use the number of publications of research in the Field of Science and Technology classification, a field-weighted citation impact, and the share of papers in the top 10 per cent of journals with regards to citations and use number of publications, field-weighted citation impact, and share of publications in top 10 per cent journals by citation impact.

[6] Aksnes *et al.* (2019) provide a comprehensive summary of the limitations of bibliometric measures of research performance and a synthesis of relevant literature. The discussion in Dale and Goldfinch (2009) illustrates some of the problems concerning the use of bibliometric indices to evaluate and compare research performance in the context of a specific discipline.

concluded there was 'positive but relatively weak correlation' between peer reviews and selected bibliometric indicators. It is evident from the review by Aksnes *et al*. (2019, pp. 6-7) that validation studies of concordance between peer review and bibliometric indicators remain a highly contested field. This suggests that, for the many PBRFSs where research performance is based on peer-review and there is a diverse range of performance that qualifies for assessment, the use of bibliometric measures of research output to assess the impact of PBRFSs is likely to be problematic.

This paper takes a different approach. It does not rely on bibliometric proxies of research performance, and it offers a different method of assessing the impact of a peer review PBRFS. The units of assessment vary among performance-based research funding systems; they may be based on the whole institution, or on a sub-set of researchers, or are group-based, and some are based on the assessment of individual researchers (Kolarz *et al.,* 2019). The present paper demonstrates how longitudinal data, derived from schemes where assessment is based on the performance of individual researchers, can be applied to gain valuable insights about responses by research institutions following the introduction of a PBRFS, and the impact on research performance.

A national performance-based funding system that satisfies Hicks's definition of a PBRFS is the New Zealand system introduced in the early 2000s.[7] Assessment is based on the performance of all eligible individual researchers within each university. The assessment method is based on peer review, which considers a range of research outputs over the previous six years, which are not amenable to measurement using bibliometric data. The process has remained the same for all assessment rounds. The individual performance results for all assessment rounds are maintained by the central administering agency. This paper exploits these features and uses anonymised longitudinal information, drawn from the assessments of all researchers who participated in the scheme from 2003 to 2018, to evaluate the contribution to changes in research quality of universities and disciplines of the entry, exit and research quality transformation of researchers.

---

[7] Hicks (2012) defines the main characteristics of these systems as follows: (i) research is assessed, (ii) research evaluation is *ex post*, (iii) research output and/or impact is evaluated, (iv) part of government allocation of university research funding depends on the outcome of the evaluation, and (v) the scheme is a national or regional system.

3. The New Zealand PBRFS: measurement of research quality and data

The New Zealand PBRFS was designed to unbundle the research component of Government funding of New Zealand tertiary education organisations, and allocate the research component based on research performance rather than the number of students.[8] Three measures are used to allocate Government funding to support research at universities and other tertiary education organisations. The largest component, the focus of this paper, is Quality Evaluation which comprised 60 per cent of the funds allocated following the 2003 and 2012 assessments and 55 per cent following the 2018 assessment.[9] The system changed the incentives facing individuals, departments and universities.

The cornerstone of the Quality Evaluation component is a process that assigns each researcher to a quality category (QC) by a complex peer-review process undertaken by a panel of experts in each subject area. These QCs are used to allocate funding to universities, and are used to compute a quantitative performance score, referred to as an Average Quality Score (*AQS*) for each subject area and university. This section defines the *AQS*, and provides their values by university and discipline for each assessment round.[10]

For each researcher portfolio submitted, the QC is determined for each individual, h, by a panel assigned to a subject area or group of subject areas.[11] There are four QCs, indicated by A, B, C and R, where the highest category is A, and R indicates an absence of significant research outputs. Each individual is given a cardinal score, G_h, depending on the letter grade: 10 for A; 6 for B; 2 for C; and 0 for R. The average quality score, *AQS*, is the employment-weighted arithmetic mean score, defined as follows.[12]

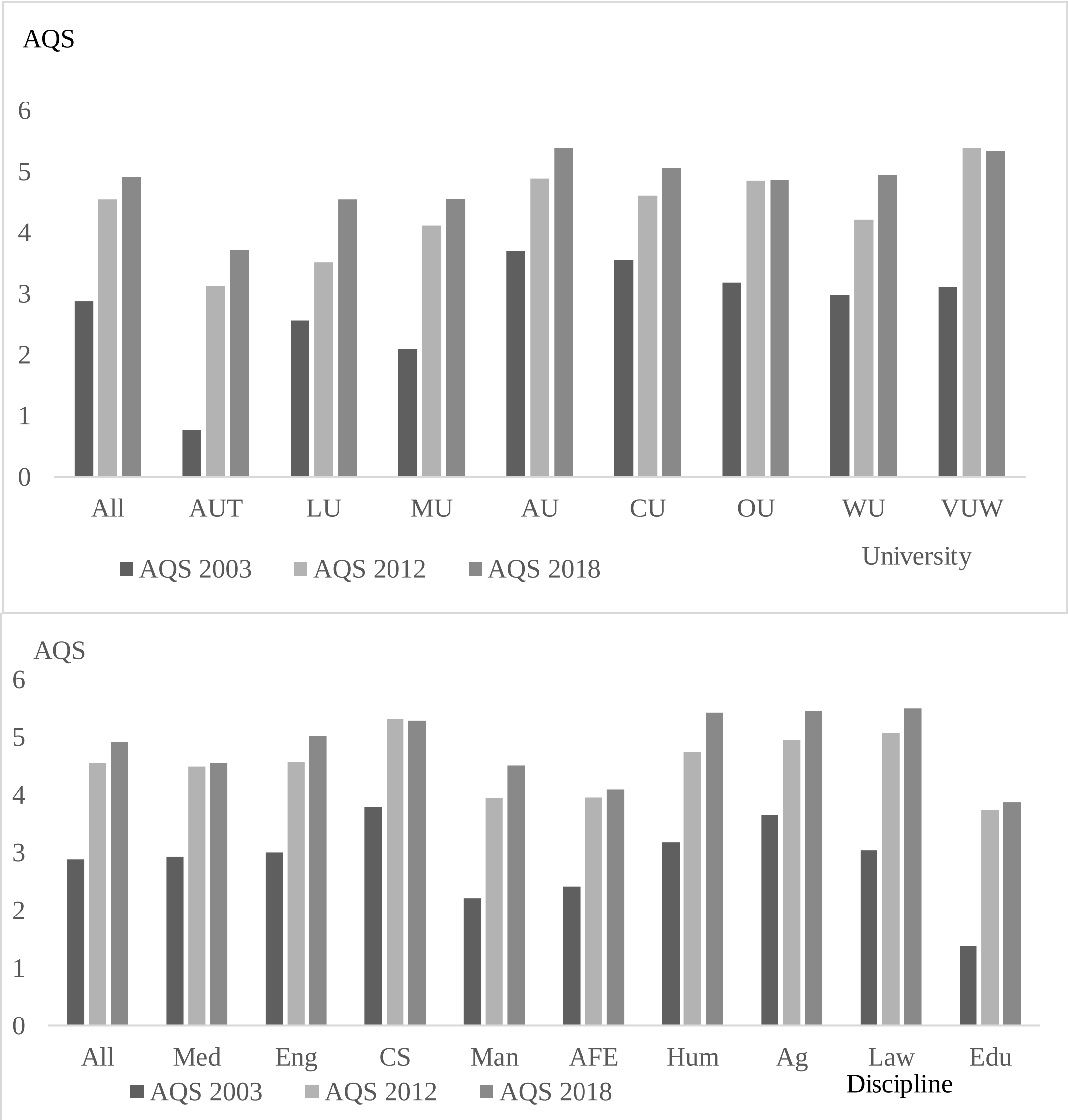
Define the full-time equivalent (FTE) employment weight of person, h, as $e_h \leq 1$, and let n denote the relevant number of employees in the group. The *AQS* is:

$$AQS = \frac{\sum_{h=1}^{n} e_h G_h}{\sum_{h=1}^{n} e_h} \quad (1)$$

[8] See Tertiary Education Commission (2019, p. 11).
[9] The other components are Research Degree Completions (25 per cent of the allocated funds) and External Research Income (15 per cent of the allocated funds following the 2003 and 2012 assessments and 20 per cent following the 2018 assessment).
[10] There was only a partial round in 2006.
[11] The assessment and scoring method used in the New Zealand PBRF system for the 2003 and 2012 rounds are described in more detail and critically evaluated in Buckle and Creedy (2019b).
[12] Two new categories C(NE) and R(NE) were introduced by the TEC for the 2012 and 2018 assessment rounds to signify if a researcher met the 'new and emerging' criteria. The score assigned by the TEC to R(NE) was 0 in 2012 and 2018, the same as in 2003. The score assigned by the TEC to C(NE) was 2 in 2012, the same as for C, and 4 in 2018 (The Tertiary Education Commission, 2019, p. 13). For the purposes of this paper the score assigned to C(NE) in 2018 is set equal to 2, to ensure consistency over time.

Since the grade for R staff is equal to zero, their number only affects the denominator in (1).

Figure 1. University and discipline AQSs in 2003, 2012 and 2018.



Notes: **Universities**: AUT = Auckland University of Technology; MU = Massey University; LU = Lincoln University: AU = Auckland University; CU = Canterbury University; OU = Otago University; WU = Waikato University; VUW = Victoria University of Wellington. **Disciplines**: Med = Medicine, Eng = Engineering; CS = Core Science; Man = Management; AFE = Accounting, Finance and Economics; Hum = Humanities; Ag = Agriculture; Law = Law; Edu = Education.

The data used here include the QC assigned to every researcher who participated in each assessment round in 2003, 2012 and 2018.[13] It also includes an anonymous identifier, age, research subject area, university of employment, and employment status (in terms of full-time

[13] The anonymised data used in this study are not publicly available and were provided by the Tertiary Education Commission (TEC) following a confidentiality agreement.

equivalent, FTE) including whether the researcher exited or entered the entire New Zealand university system between rounds or transferred to or from another NZ university.

The *AQS*s for each university and discipline group are shown in Figure 1.[14] The top panel of Figure 1 shows the *AQS*s for each university in each assessment round, while the bottom panel shows the equivalent for each discipline group. For the entire university system, there was a rise in average research quality, as measured by the *AQS*. The FTE-weighted average *AQS* increased between 2003 and 2012 at an annual compound growth rate of 5.2 per cent. For the six-year period 2012 to 2018 this rose at a much lower annual compound rate, of about 1.3 per cent. This broad pattern of higher initial growth in research quality followed by slower subsequent growth holds for all but one university (Lincoln) and all disciplines. If the initial growth in average research quality can be, at least in part, attributed to the introduction of the PBRFS, the impact on *AQS* growth appears to be diminishing. The suggestion of a fading effect is consistent with the findings of several other studies; see the review by Checchi *et al*. (2019).

4. Effects on university incentives of PBRFS

The growth in average research quality results from a complex dynamic process of researcher transitions, which determine the composition and quality of researchers within universities and across disciplines. Consistent with the Schumpeterian analogy, the turnover of staff in the NZ university system as a whole was high: over the period 2003 to 2012 there was a turnover rate of 53 per cent, and over the whole period 2003 to 2018 this rate was 81 per cent.[15]

The PBRFS introduced new financial incentives for the management of New Zealand universities, by directly aligning Government funding of research with the perceived quality of researchers. The new system of funding also introduced new reputational incentives and new information about the relative research quality of universities and disciplines. The results of the first assessment round in 2003 provided all universities, for the first time, with publicly available information on what was regarded as high-quality research, and how this was distributed among universities and disciplines.[16] It created the type of knowledge environment

[14] The subject composition of each discipline group is explained in Buckle and Creedy (2020).
[15] The turnover rate is measured as the number of researchers who exited, divided by half the sum of the initial and final total number of FTE researchers.
[16] See the information provided by the publication of the results from the first assessment in 2003 in The Tertiary Education Commission (2004).

and financial incentives that could be expected to have triggered a process comparable to Schumpeter's 'gale of creative destruction' (Schumpeter, 1942, pp. 81-86).

Schumpeter's 'gale of creative destruction' was in reference to new products and production technologies changing relative prices and driving old products and technologies from the market. Although the process of change in research quality within universities may be qualitatively different, it involves responses to changes in relative prices, prompted by the PBRFS, that alter the research revenue distributed to universities in a manner that is directly associated with the research calibre of their researchers. These relative price changes lead to processes that alter the quality distribution of researchers. To paraphrase Schumpeter's description of the process of industrial development, the new incentives intensified a process of mutation of university research quality whereby the university structure was reformed from within, incessantly improving or removing researchers, incessantly recruiting new ones.

4.1 Financial incentives and reputational effects

The introduction of the PBRFS in the early 2000s meant that universities were incentivised to recruit higher-quality researchers, as measured by their *AQS*. These incentives were both reputational and financial.[17] The new funding system directly linked a dollar amount of research funding to each university per 'point' in the numerical score, G_h, where $h=1,\dots,n$, and n is the number of researchers. Thus, for example, the research funds received by a university for an A-rated researcher (for whom $G_h = 10$) was five times the amount received for a C-rated researcher ($G_h = 2$). Therefore, following the first assessment round in 2003 there was a positive proportional relationship between future research funding to each university and its aggregate score: $\sum_{h=1}^{n} G_h$.

However, the incentives vary among universities depending on their initial *AQS*. For example, the proportional increase in the *AQS* of a university with n researchers (assuming for this illustration these are all full-time employees), $\frac{\Delta AQS}{AQS}$, resulting from hiring one additional type-A researcher, is given by:

$$\frac{\Delta AQS}{AQS} = \frac{\left(\frac{w_A}{AQS} - 1\right)}{n+1} \qquad (2)$$

[17] Results of each assessment round are published by The Tertiary Education Commission (2004, 2013, 2019).

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

$$\frac{\Delta AQS}{AQS} = \frac{n_R}{n - n_R} \quad (3)$$

Figure 2. Effects on AQS of varying Quality Categories

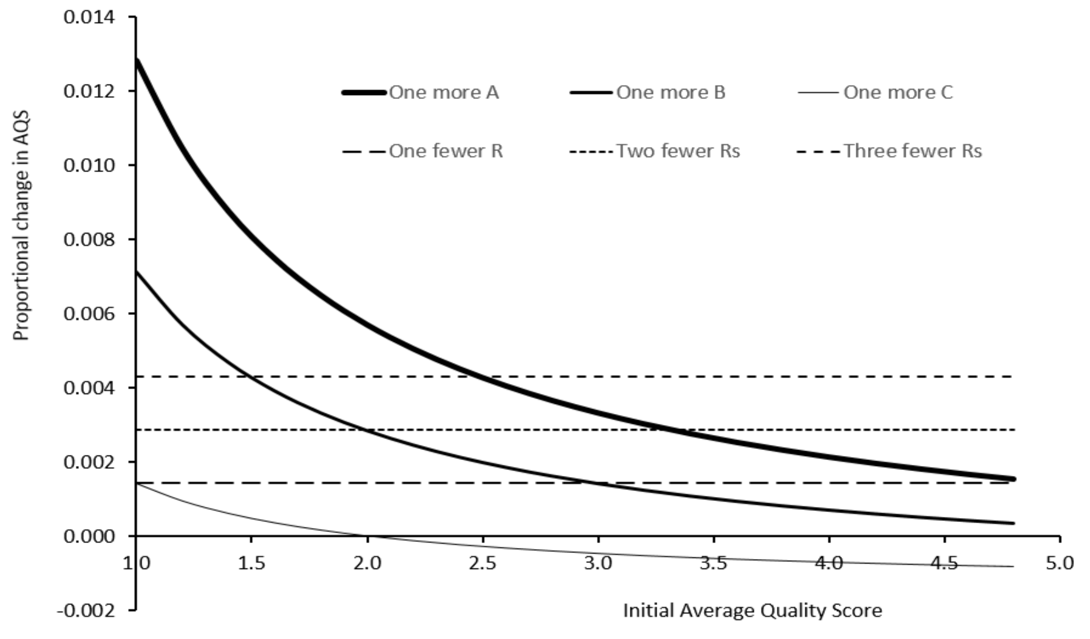


Figure 2 illustrates some variations in the change in *AQS* for the case where $n = 700$. Clearly, if a university has an *AQS* greater than 2, its *AQS* and revenue per researcher would fall as a result of hiring one more C-type researcher. It would need to recruit at least at level B, or reduce the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

4.2 Hypotheses regarding organisational change

Faced with a strong incentive to improve the research quality of staff, several organisational responses are possible, 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, there is some evidence that management practices have responded to the changed incentives created by the introduction of a PBRFS.[18] It is therefore useful to consider the broad characteristics that may be expected to follow from an institutional objective of maximising, subject to constraints, the *AQS*. The following considers the expected responses of exits, entrants, and quality transformations in turn. In each case the hypothesis is stated, followed by its rationale.

H1 *(Exits): The rate of exit of Rs is higher than for other categories for 2003-2012.*

Since $G_h = 0$ for Rs, and therefore no funding is available through the PBRFS, an obvious method of increasing the *AQS* is to minimise the number of R-quality staff. This process would require resourcing, and given the multiple outputs of a university, involving teaching and external engagement as well as research, an appropriate staffing balance is obviously required. Nevertheless, organisational responses which reduced the proportion of R researchers would enhance all university and discipline *AQSs*.

H2 *(Exits): The rate of exit of Cs for 2012-2018 is higher than that of As and Bs.*

In 2012 the *AQS* for most universities and disciplines exceeded the G_h for Cs. While the same staffing balance and cost constraint issues are relevant as for *H1*, the option of reducing the proportion of C's is consistent with increasing the *AQS* for all universities.

H3 *(Entrants): The entry rate of As, Bs and Cs for 2003-2012 is higher than that of Rs.*

Given the $G_h = 0$ for Rs, for the same reasons motivating *H1*, a low rate of recruitment of Rs compared to As, Bs and Cs is expected. Although the university-wide *AQS* in 2003 > G_h for Cs, several universities and disciplines had *AQS*s below 2 and the possibility of substitution of Cs for Rs, coupled with the cost of Cs compared to As and Bs in the same discipline, implies that a high rate of recruitment of Cs is likely in the 2003 to 2012 period.

H4 *(Entrants): The entry rate of As and Bs during 2012-2018 is higher than during 2003-2012.*

[18] Woerlert and McKenzie (2018) found that Australian universities tended to replicate the national performance indicators used in the national PBRFS in their individual-level performance management frameworks for academic staff.

The higher G_h for As and Bs compared to Cs, and an increase in the *AQS* beyond the G_h for Cs for most universities by 2012, implies that a rise in the entry rate for As and Bs, compared with that for Cs (and Rs), is likely for the 2012-2018 period.

H5 *(Transformations): The rate of upward transformation of Bs and Cs to As is higher than Rs.*

The time and resources required to convert an R researcher into an A quality researcher is likely to be longer and more expensive than to convert a B or C researcher to an A. Similarly, the transformation rate of Bs to As is expected to be higher than for Cs to As.

H6 *(Transformations): The proportion of A plus B researchers in 2018 is higher than in 2003.*

If hypotheses *H1* to *H5* hold, by 2012 there is likely to be a higher proportion of researchers in QC categories with $G_h > AQS$. The all-universities *AQS* in 2012 was 4.55, and hence it is expected that the sum of A and B researchers as a proportion of total researchers increased by 2018 (compared to 2003 and 2012).

H7 *(Transition dynamics): There are differences in transition patterns among disciplines.*

Differences among disciplines in the level and rate of improvement of research quality could arise from differences in research methods, funding opportunities, alternative labour market opportunities, and the opportunity costs of careers in academia and research. For example, Xu (2008) found substantial and systematic variations among disciplines with respect to the main factors that determine turnover. Salaries and alternative career opportunities influence the mobility of staff and the ability of institutions to recruit staff in the education field; see for example Ehrenberg *et al.* (1991). Boyle (2008) demonstrated that differences in labour market opportunities affected the relative research performance of disciplines in the first New Zealand round.

Similarly, individual characteristics (age, experience, gender), institutional characteristics (performance management, commercial orientation, the presence of PhD programmes, and institutional mission and priorities), research methods (collaboration, team research, joint authorship), and the availability of contestable research funding, can influence research productivity, and these all vary among disciplines. For example, Wanner *et al*. (1981) found that discipline influenced the research environment and type of research output. Shin and Cummings (2010) concluded that staff collaboration, time spent on research, institutional goal-orientation and institutional mission influenced academic publishing, and these influences differed by discipline. Jung (2012) concluded, for Hong Kong academics, that the influence of individual and institutional factors on self-reported research output differed significantly among natural sciences, engineering, and medical science disciplines and the humanities, social sciences, and business disciplines. Sabharwal (2013) found that research productivity was associated with discipline and career stage.

5. Mapping researcher dynamics and the 'gale of creative destruction'

The availability of longitudinal data makes it possible to reveal changes for each researcher in the system, and in turn, evaluate how these dynamics changed the *AQS* of the entire university system, individual universities and disciplines. The primary tool for the analysis of flows is the social accounting framework, with its core component, the transition matrix, showing the movements of individuals (or full-time-equivalents) between discrete categories, or 'states', from one period to another.[19] The transition matrix of flows reveals in a transparent manner the predominant changes that lie behind the *AQS* increases from one assessment round to another. This section focusses on the dynamics for the entire university system.

The transition proportions (the flows as a proportion of the initial total number in each category) for all universities combined are shown in Table 1, for movements from 2003 to 2012, and from 2012 to 2018. The flows are from rows to columns. Consistent with *H1*, the largest exit rate during the period 2003 to 2012, at just over 70 per cent, is of those assessed as R-quality researchers in 2003.

Table 1 shows that, consistent with *H2*, the exit rate of Cs during the period 2012 to 2018, at just over 39 per cent, was higher than for As (27.2 per cent) and Bs (29.9 per cent). The total number of exits of Cs was just under 48 per cent of total exits. A higher entry rate for As, Bs, and Cs in 2003 to 2012 compared to that for Rs is consistent with *H3*. The highest entry rate is for Cs, at nearly 52 per cent. This is consistent with the change in reputational and financial incentives faced by universities and the need to substitute new researchers for the high rate of exit of Rs that occurred during 2003 to 2012. However, contrary to *H4*, there is no evidence of a tendency for the entry rate of As and Bs to rise during 2012 to 2018 compared to the

[19] It seems that the first papers applying the social accounting framework to the education sector were by Stone (1965, 1966). This work involved the application of input-output analysis and Markov chain methods for the purpose of calculating resources, including demographic, required to achieve specified educational outputs.

entry rate for Cs, which remained the dominant source of new entrants during 2012 to 2018. This may well reflect the relative scarcity of higher-quality researchers.

Table 1. Matrix of flows: all universities 2003 to 2012 and 2012 to 2018

	Category in 2012					
Category in 2003	A	B	C	R	Exits	Total
A	**0.531**	0.151	0.005		0.313	423.55
B	0.178	**0.373**	0.086	0.001	0.362	1689.16
C	0.029	0.254	**0.223**	0.011	0.483	2212.23
R	0.003	0.061	0.187	**0.046**	0.702	2206.19
Entrants	0.076	0.352	0.515	0.057		3079.71
Total	832.13	2475.42	2639.16	303.99	3360.14	9610.84
	Category in 2018					
Category in 2012	A	B	C	R	Exits	Total
A	**0.600**	0.125	0.005		0.272	832.13
B	0.180	**0.440**	0.079	0.002	0.299	2475.42
C	0.018	0.293	**0.285**	0.012	0.392	2639.16
R		0.078	0.321	**0.062**	0.539	303.99
Entrants	0.057	0.304	0.596	0.043		2969.33
Total	1158.62	2894.20	2818.85	182.91	2165.45	9220.03

Note: The sum of category A, B, C, R portfolios in the 'Total' rows and columns in each table equal the respective 2003, 2012 and 2018 portfolio totals.

The quality transformation rates of B, C and R into A researchers is consistent with *H5* in both periods. The transformation rate for Bs to As was about 18 per cent in both periods; for Cs it was close to 3 per cent in 2003 to 2012 and 2 per cent in 2012 to 2018. For Rs, the transformation rate was almost zero in 2003 to 2012 and was zero in 2012 to 2018. The same pattern is evident for the rates of quality transformation of Cs and Rs to Bs and as expected, the transformation rate of Cs and Rs increases as the target QC is reduced.

The patterns of change suggested by the hypotheses discussed above suggest differences in the observed frequencies between the two periods, as suggested by hypotheses *H1 to H6*. This is confirmed by the χ^2 statistic, calculated by comparing the frequency distribution observed for 2012 to 2018 with an expected frequency distribution based on the observed frequencies for 2003 to 2012. An adjustment was made to allow for the fact that the 2003 to 2012 frequency distribution is for a nine-year period, whereas 2012 to 2018 is a six-year period.

The χ^2 statistic is equal to 992, which compares with the critical value (for a type I error of 0.01, and for 16 degrees of freedom) of 32.00, confirming a significantly different pattern of transition dynamics during the second period.[20] The main sources of the difference are: a higher exit rate for all grades; a higher proportion of entrants of Cs and Rs; a tendency for a higher upward transformation rate for Bs, Cs and Rs; and a smaller proportion remaining in the same QC during 2012 to 2018.

A large change involves the number of Rs over the period. The number of Rs, as a proportion of their initial stock, fell from 34 per cent in 2003, to about 5 per cent in 2012, and less than 3 per cent in 2018, with exit rates over the two periods of 70 and 54 per cent. The number of As and Bs, as proportions of their initial stocks, rose over both periods, supporting hypothesis *H1*.

The same approach can be applied to individual universities, to examine the extent of heterogeneity among universities (for all discipline groups combined) in their researcher dynamics. Given space constraints, the tables are not reproduced here. To consider, for both periods, whether individual universities differ from the pattern revealed by all universities combined, the flows for each university which would result from applying the transition proportions and exit rates for all universities were computed. Using these hypothetical flows as 'expected' values, a χ^2 test can be carried out. Importantly, the universities were found to differ from each other in ways that are consistent with the hypotheses in Section 4. Furthermore, there was a tendency for those universities with initial relatively high *AQS*s also to have higher rates of transformation or conversion of their initial stock of researchers to higher QCs than for those universities having lower initial *AQS*s. This is consistent with the research findings discussed in Section 4.1, which identified institutional factors such as goal orientation and institutional mission as having an influence on research performance (Shin and Cummings, 2010; Jung, 2012).

Analysis of discipline groups reveals greater diversity of transition rates than among universities, in both periods. During 2003 to 2012 only Management had transition rates that do not differ significantly from those of all disciplines combined. The discipline with the largest χ^2 value is Education, which had the lowest *AQS* in 2003, the largest change in total researchers, and the highest growth rate of *AQS* during the period. During 2012 to 2018

[20] The χ^2 statistic = 315 when expected 2012 to 2018 frequencies are based on actual 2003 to 2012 frequencies.

Management, Core Science and Engineering had transition dynamics not significantly different from those for all disciplines combined. These results are consistent with hypothesis, *H*7.

6. Decomposing contributions of entry, exit and quality transformation to AQS growth

Further insight into the nature of the differences among universities and discipline groups in their dynamics can be obtained by decomposing the *AQS* changes into component contributions. The decomposition method suitable for this purpose recognises that to improve the research quality of an institution, there are several ways in which it can transform the quality of its stock of researchers. These include retaining good researchers, creating opportunities to improve the research skills of those researchers who remain within the institution, 'managing out' the lower-quality researchers, and recruiting new researchers of a higher quality than incumbent or exiting researchers.

Institutions do not have complete control over all these changes. Competition for research skills, contractual and legal obligations, and individuals' career choices and retirement decisions impinge on the extent to which these transitions can be managed by an institution. Hence, many of the observed flows may be the result of other forces, but at the margin the institution can be expected to have an influence through human resource management policies such as recruitment decisions, performance management, training and support of a strong research environment.

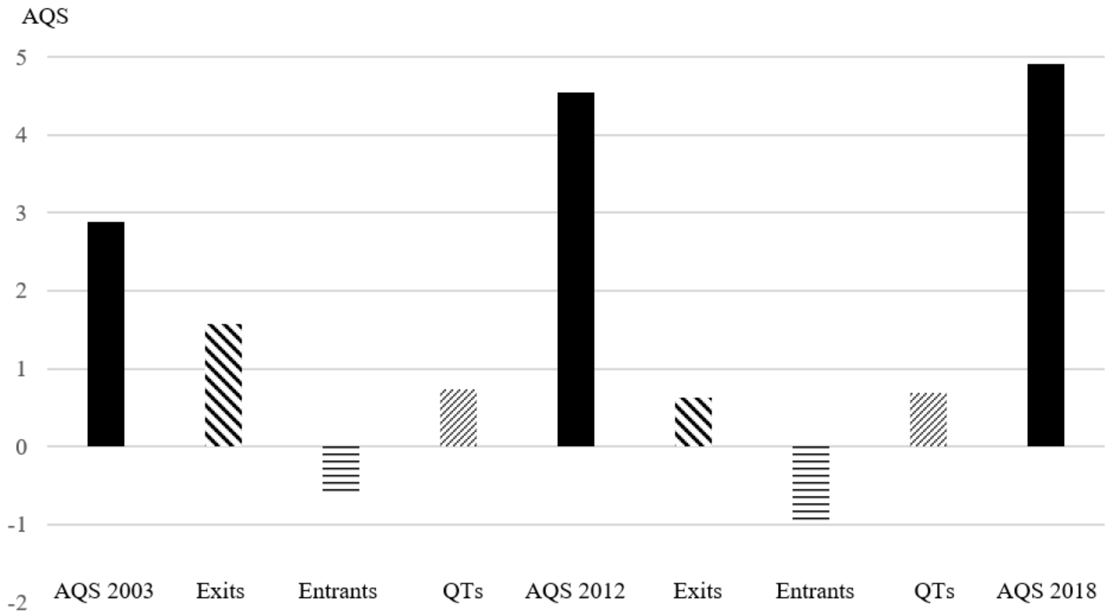
The required decomposition can be obtained by suitably modifying the relevant transition matrix, as follows. Let the initial and final *AQS* for the specified group be denoted by *Q*1 and *Q*2. The aim is thus to decompose *Q*2-*Q*1 into components that measure the separate impact on the change in *AQS* of exits, entrants and quality transformations. An *AQS* can be calculated for the final period, using an assumption that there are no entrants into any category: this is denoted by *Q*3. Furthermore, it is possible to obtain an *AQS* in the final period, by setting all exits and entrants to zero: this is denoted *Q*4. The assumption of no exits is applied by supposing that those recorded as exiting remain in the quality category in which they were placed in the initial period: that is, the diagonals of the flow matrix are augmented by the number of (FTE) exits. The difference, *Q*4-*Q*1, therefore reflects the effect of quality transformations made by those who remain in the system. The difference, *Q*3-*Q*4, reflects the separate effect of exits. Finally, the difference, *Q*2-*Q*3, measures the effect of entrants. These are combined to give:[21]

$$Q2 - Q1 = (Q2 - Q3) + (Q3 - Q4) + (Q4 - Q1) \tag{4}$$

These contributions for all universities combined are summarised by Figure 3. The various components, for each university (all disciplines) and discipline group (all universities) can be obtained in the same way, revealing that the net impact of exits on *AQS*s is positive for all universities and disciplines in both periods; the net impact of entrants is always negative; and the net impact of quality transformations is always positive.

For all universities, three results stand out from Figure 3. First, the positive contribution from exits was much larger in 2003 to 2012 than the subsequent period: it fell from 1.57 to 0.63. Second, the negative contribution from entrants was larger in the second period: it increased from -0.63 to -0.96. Third, the contribution from the quality transformation of researchers was similar in both periods (at 0.73 and 0.69 in the respective periods).

Figure 3. Contributions to changes in all-universities AQS of exits, entrants & quality transformations



[21] The alternative expression, $Q2 - Q1 = (Q2 - Q4) + (Q4 - Q3) + (Q3 - Q1)$, does not have an interpretation that is convenient for the purposes of the analysis in this section, because it does not identify the effect of quality transformations on changes in an *AQS*.

During the first period the largest proportion of exits was of Rs. During the second period, the largest proportion of exits were of Cs and there were more B and A exits than during 2003 to 2012. This is unsurprising because, as the quality of the stock of researchers improves, the quality of exiting researchers will tend to rise, for a given turnover rate. The exit pattern is consistent with the incentives described in Section 4, especially hypotheses *H1* and *H2*. Using the same decomposition for individual universities, it was found that there is a large range in the impact of exits during 2003 to 2012.

The effect of entrants on *AQS* is consistently negative in both periods, implying that the average research quality of entrants was lower than the *AQS* of the initial stock of researchers. This is the case in both periods, but is higher in the second period because the higher initial *AQS* (in 2012) makes it more difficult to recruit entrants above the average quality. The net impact of quality transformations is, not surprisingly, consistently positive. This reflects decisions to attract and retain good researchers who are likely to improve over time. Nevertheless, this is the average effect of those who improve and those who decline in quality, while remaining within the institution. This pattern is similar in both periods and supports *H5*.

There are differences between universities and disciplines in the range of the effects of exits, entrants and quality transformations. For universities, the largest range in both periods is in the contribution of exits to improved *AQS*s. For disciplines, the largest range in both periods is in the contributions of entrants to improved discipline *AQS*s. The range of contributions from quality transformations are smaller than for exits and entrants in both periods and the ranges, though slightly higher for disciplines, are smaller both for universities and disciplines.

7. Contributions of researcher quality and discipline composition to *AQS* changes

One further way for a university to improve its *AQS*, not so far discussed in detail, is to exploit the systematic differences among discipline groups (as shown in Figure 1) by changing the discipline composition of its researchers.[22] Changes in the overall discipline shares between 2003 and 2018 are small. Nevertheless, these do in some cases appear to be related to the initial *AQS.* For example, Education had the lowest initial *AQS* and the largest

[22] Changes in the discipline composition of staff, following the introduction of performance measures by many states in the USA focussing on the evaluation of teaching, are reported by Payne and Roberts (2010).

reduction in the share of researchers.[23] The next lowest initial *AQS* was for Management, which also experienced a reduction in the overall share of researchers. Agriculture and Engineering had relatively high initial *AQS*s and saw increased shares over the period. By contrast, Medicine had the largest share increase but a relatively low initial *AQS,* and Core Science, with a reduction in its share, had a relatively high initial average quality.

The present section is thus concerned with the question of the extent to which *AQS* changes can be attributed to the changing discipline composition of its staff compared with the changing quality of its researchers. The decomposition method used for this purpose follows the approach suggested by Shorrocks (2013), which was initially used to examine the contribution of different factors (such as the population structure and the tax system) to changes in income inequality over time. Consider a single university, with *D* disciplines. The proportion of staff in discipline, *j*, at time, *t*, is denoted by $p_{j,t}$. Let the *AQS* in discipline, *j*, at time, *t*, be $q_{j,t}$. Changes in $q_{j,t}$ reflect the combined effect of turnover and changes experienced by staff who remain in the same organisation. However, the following is concerned only with changes in the research quality of the discipline, $q_{j,t}$. The *AQS* of a university, Q_t in period, *t*, is equal to:

$$Q_t = \sum_{j=1}^{D} p_{j,t} q_{j,t} \tag{5}$$

In the following, it is convenient to write $Q_t = Q\left(q_t, p_t\right)$, where q_t and p_t are the vectors, $\left[q_{1,t},...,q_{D,t}\right]$ and $\left[p_{1,t},...,p_{D,t}\right]$. The change in quality, ΔQ, from period *0* to period *1*, is given thus $\Delta Q = Q_1 - Q_0 = Q\left(q_1, p_1\right) - Q\left(q_0, p_0\right)$. This can be expressed as:

$$\Delta Q_A = \left[Q\left(q_1, p_1\right) - Q\left(q_0, p_1\right)\right] + \left[Q\left(q_0, p_1\right) - Q\left(q_0, p_0\right)\right] \tag{6}$$

The first term reflects the change in *AQS* attributed to the changing quality of its staff, given the discipline composition in period *1*. The second term reflects the change in *AQS* attributed to the changing discipline composition of the university, given the quality of staff in period 0.

However, the change in *AQS* may also be decomposed as:

[23] Education numbers were clearly affected, in the early years of the PBRFS, by substantial legislated institutional changes involving the merger of colleges of education with universities.

$$\Delta Q_B = \left[Q(q_1, p_0) - Q(q_0, p_0) \right] + \left[Q(q_1, p_1) - Q(q_1, p_0) \right] \quad (7)$$

The first term reflects the change in *AQS* attributed to the changing quality of its staff, given the discipline composition in period 0. The second term reflects the change in *AQS* attributed to the changing discipline composition of the university, given the quality of staff in period *1*. There is no special presumption in favour of using (6) or (7). Hence, one approach is to take the arithmetic mean of the components.[24]

Table 2. Quality and composition change contributions to improvements to each university's research quality

University	Increase in *AQS* 2003-2012	Attributed to: Quality	Composition	Increase in *AQS* 2012-2018	Attributed to: Quality	Composition
All	**1.67**	1.67	0.00	**0.36**	0.37	-0.01
AUT	**2.38**	2.36	0.02	**0.58**	0.60	-0.02
Lincoln	**0.95**	0.95	0.00	**1.03**	1.03	0.00
Massey	**2.01**	1.95	0.06	**0.45**	0.44	0.01
Auckland	**1.22**	1.17	0.05	**0.49**	0.49	0.00
Canterbury	**1.06**	1.03	0.03	**0.44**	0.38	0.06
Otago	**1.67**	1.65	0.02	**0.01**	-0.03	0.04
Waikato	**1.23**	1.14	0.09	**0.61**	0.58	0.03
VUW	**2.27**	2.22	0.05	**-0.04**	-0.10	0.06

Notes: This table shows the absolute changes to a university *AQS* attributed to changes in research quality and discipline composition. The proportions attributed to these two components can be calculated by dividing the amount attributed by the increase in *AQS.*

The decompositions are shown in Table 2, which reports absolute changes to the *AQS* attributed to changes in research quality and discipline composition (in FTEs) for each university. During 2003 to 2012, all the increase in the *AQS*, for all universities combined, came from quality changes. Among universities, there are some small contributions from discipline composition change, though in one case it is negative. During 2012 to 2018 the contribution from quality improvement was substantially smaller than in the previous period, but was still the dominant source of improvement in *AQS*. Hence, NZ universities generally

[24] In the earlier analysis of the contributions of exits, entrants and transformations to the change in *AQSs*, the perspective unambiguously concerns movement from an earlier to a later period, so only one expression of the decomposition is relevant. The present context involves changes arising from differences in the composition of the population.

(the main exception being for education) did not exploit strategies involving reductions in the size of academic departments. There are of course teaching commitments, and other strong constraints which would prevent senior managers from imposing substantial variations in staff/student ratios. Using the Schumpeterian analogy, despite institutional constraints and quasi-independent sections that make up universities, the 'creative destruction' has involved a high turnover of staff within academic disciplines and, to a smaller extent, the 'transformation' of researchers, rather than the 'destruction' of disciplines within universities

8. Assessing sustainability of researcher dynamics

The approach used above has involved the formulation of a set of hypotheses about the way in which university employment dynamics may be expected to change in response to a new set of incentives created by a PBRFS. The various analyses of transition matrices, facilitated by access to anonymous longitudinal data on individual researchers, have largely supported the hypotheses. The results therefore support the argument that, despite the necessary lack of comparable data for the pre-PBRFS, the new incentives did have a substantial effect on NZ universities.

There is no doubt that New Zealand universities in 2003 had a large proportion of R researchers, as shown in Table 1, and that it was inconsistent with the incentive structure created by the PBRFS. Indeed, the subsequent high exit rate and low recruitment rate of Rs over 2003 to 2012 illustrates this point. If, prior to 2003, universities were satisfied with the quality distribution of researchers, the post-PBRFS environment changed what they considered to be a desirable distribution. Moreover, once the stock of R researchers had fallen to only 5 per cent by 2012, further adjustments to the quality distribution of researchers required a different response in the subsequent 2012 to 2018 period.

It has been seen that the period 2012 to 2018 did indeed experience a different pattern of transition dynamics, one that resulted in a sustained low level of Rs coupled with a growing share of A and B researchers. This was achieved by a higher retention of As and Bs and a higher positive quality transformation rate. The differences in the transition dynamics between the first and second periods suggests a sequencing of priorities in the way universities managed to improve the average research quality of researchers. This involved a process of entry and exit that resulted in the initially relatively high proportion of lower-quality researchers being replaced by higher-quality researchers, and greater emphasis on retaining higher quality researchers.

The present section presents a further approach, involving the use of the observed dynamics to generate an equilibrium stock of researchers. In equilibrium, the exits and entries, combined with transitions among quality categories, are such that the distribution of researchers across QCs remains constant. This equilibrium distribution can be compared with the actual stocks observed at the end of each assessment period. A large disparity and an unrealistic equilibrium distribution suggest that the changes observed since the PBRFS was introduced cannot be sustained indefinitely. Hence, it is reasonable to argue that the changes are sufficiently large to have been induced by the new incentives.

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 fixed transition matrix. Then for a constant entry vector, *b*:

$$n_t = b + Tn_{t-1} = b + T\left(b + Tn_{t-2}\right) \tag{8}$$

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:

$$n = \left(1 + T + T^2 + T^3 + ...\right)b = (I - T)^{-1}b \tag{9}$$

Using the transition matrix and vector of births in Table 1 gives equilibrium stocks of A, B, C and R researchers respectively for each of the periods 2003 to 2012 and 2012 to 2018. The results for all universities combined are shown in Table 3. The equilibrium total stocks for the 2003 to 2012 transitions exceed actual stocks in 2012 by about 23 per cent. It is therefore reasonable to suggest the transition dynamics generating the evolution of the quality of researchers at New Zealand universities after 2003 represents to some extent a structural shift in response to the introduction of the PBRFS, and this path cannot be expected to continue indefinitely. While equilibrium stocks exceed actual stocks for each QC in 2012, the largest differences are in the stocks of R followed by A researchers.

The equilibrium total stocks for the 2012 to 2018 transitions exceed actual stocks in 2018 by about 26 per cent. Although the nature of the dynamics changed during the second period, they continued to generate large differences between the actual and simulated equilibrium stocks. This is also illustrated by the differences between the actual *AQS*s in 2012 and 2018 and the implied equilibrium *AQS*s. The value of the implied equilibrium *AQS*s are 4.46 for the 2003 to 2012 dynamics and 3.90 for the 2012 to 2018 dynamics. The actual values are 4.55 for 2012 and 4.91 in 2018. Both values are in excess of the equilibrium values, by about 2 per cent in 2012 and 26 per cent in 2018. The *AQS*-increasing strategy depends on the initial *AQS*, so that to continue with fixed transition proportions is not optimal.

Table 3. Initial and equilibrium portfolios and QC distributions: all universities combined

QC	2003	2012	Equilibrium 2003-12	% Diff	2018	Equilibrium 2012-18	% Diff
A	424	832	1339	60.9	1159	1248	7.7
B	1689	2475	2511	1.5	2894	2522	-12.9
C	2212	2639	2928	11.0	2819	3566	26.5
R	2206	304	923	303.6	183	1567	756.3
Totals	6531	6250	7701	23.2	7055	8903	26.2

Notes: Actual portfolios for each QC and the transition frequencies and entrants applied to expression (9) to derive equilibrium distributions of portfolios by QC are derived from Table 1. %Diff is the percentage difference between the equilibrium stocks and actual stocks where %Diff = 100*{[(Equilibrium stock-Actual stock in t+1)/Actual stock in t+1]-1].

Applying the same approach to individual universities and disciplines, the results (not shown in view of space constraints) show large variations in the extent to which the equilibrium stocks of researchers (generated by both 2003 to 2012 and 2012 to 2018 transitions and entries) exceed actual stocks. The results provide a stark indication of the large extent of the changes in the distribution of the quality of researchers since the introduction of the PBRFS. This is despite the many characteristics of universities which make structural change difficult. This provides indirect support for the idea that the observed changes in the distribution of the quality of researchers since the introduction of the PBRFS do represent, to some extent, the influence of the introduction of the new research funding system.

9. Conclusions

The aim of this paper was to explore methods which can be used to evaluate the effects on a measure of university research quality of the introduction and continued use of a PBRFS. This presents a special challenge because consistent measures of research quality are typically not available before a PBRFS is enforced. It is argued that analyses need to go beyond the collection and examination of times series of proxy measures (obtained, for example, from bibliometric sources). Any PBRFS exercise and its associated metrics of research output and quality necessarily impose a set of incentives, facing individual researchers and university managers. And of course, all participants necessarily face a range of constraints.

This paper has suggested that, without attempting to develop a full-blown model of a university or research unit, containing the structural specification of multiple objectives, cost functions and organisational constraints, it is possible to form a set of hypotheses about the likely responses to incentives created by the specific design features of a PBRFS. The hypotheses can then be used to motivate and inform empirical analyses and tests.

The analyses presented here are particularly appropriate for those PBRFS which are based initially on the evaluation of individual researchers, along with the use of an aggregation procedure to measure the research quality of an organisation, or unit within the organisation, or even of all research organisations within a country. Furthermore, the methods rely on the use of (anonymised) longitudinal data, collected over a period containing two or more measurement ‘rounds’. The fundamental framework of analysis is the ‘social accounting matrix’, which traces entries into, exits from, and transitions within, a research unit over a given time period. In examining these processes by which research units are transformed over time, an analogy was drawn here with the ‘gale of creative destruction’ famously described by Schumpeter in the context of the ‘births’ and ‘deaths’ of firms and economic growth: here perceived higher-quality researchers replace lower-quality researchers in a complex process of organisational change.

The methods were applied to a special dataset covering the period, 2003 to 2018, during which the New Zealand PBRFS has been operating. However, the methods can be applied to any PBRFS based on individual evaluation and for which longitudinal data are recorded. In the present context, the data cover the complete population of eligible New Zealand university researchers.

Examination of the incentives created by the New Zealand PBRFS, and its special metrics (the Quality Category, QC, for individuals, and the Average Quality Score, *AQS*, for research units), suggested that the dynamics of universities and disciplines are likely to depend in particular ways on the initial structure of research units. A clear example is the fact that a unit with an *AQS* of, say 3, necessarily lowers its *AQS* by employing an extra researcher who is perceived to be assigned to a QC of C, since this has a score of only 2. The was substantial heterogeneity among universities and discipline groups, and differences in the pattern of transitions over time, that are consistent with the hypotheses derived from the special nature of incentives created by the metric used. In addition, the extent of heterogeneity was found to be lower during the later PBRFS period, 2012 to 2018.

The paper has also shown how changes in the quality measure of a research unit over time can be decomposed into changes arising separately from the patterns of entries, exits and transformations of individuals over time. Furthermore, quality changes were decomposed into contributions arising from changes in the discipline composition within a university, and changes in research quality of the university's disciplines groups. Generally, New Zealand universities were found not to have used a strategy of substantially changing their discipline composition (that is, of substituting higher-quality discipline groups for lower-quality, as measured by the *AQS*). However, during the first period 2003 to 2012, all universities substantially reduced the number of researchers in the poor-performing Education discipline group.

The sustainability of changes was examined by considering, for each university, the structure that would arise from holding the transition proportions and the entries constant over time, until an equilibrium structure is reached, where entries just compensate exits from each Quality Category. The result that the equilibrium stocks of researchers does not appear to be sustainable lends further support for the argument that the changes made over the PBRFS period were directly influenced by the particular incentives, operating in the context of an initial position.

It is suggested that the methods used in this paper enable a deeper understanding of institutional responses to Performance-Based Research Funding Systems, the sustainability of changes in research quality, and the contributions of changes in researcher quality and discipline composition to changes in institutional performance. The approach, involving explicit consideration of the special incentives for structural change which face each type of research unit, contrasts with the time-series analysis of proxy measures of research quality. Such proxy measures are unable to capture the complexities of the metrics used in practice.

The social accounting framework adopted here is clearly applicable to situations in which a PBRFS is based on the evaluation of individual researchers' quality metrics, since these are necessarily recorded in the evaluation process. Nevertheless, even where research units are evaluated as a whole, the metrics used clearly create specific individual incentives and stimulate organisational and management changes. Hence, even where individuals are not explicitly assessed in the evaluation procedure, they will inevitably be assessed by current and potential employing units in their recruitment and promotion procedures. Hence, it would be of value, in such group-evaluation processes, to collect the kind of information relating to individuals that enables the methods used here to be applied.

References

Adams, J. & K. Gurney (2010) *Funding Selectivity, Concentration and Excellence – How Good Is the UK's Research?* Oxford: Higher Education Research Institute.

Aksnes, D.W. & R.E. Taxt (2004) Peer reviews and bibliometric indicators: a comparative study at a Norwegian university. *Research Evaluation*, 13: 33-41.

Aksnes, D.W., L. Langfedlt, & P. Wouters (2019) Citations, Citation Indicators, and Research Quality: An Overview of Basic Concepts and Theories. *Sage Open*, 7th February.

Ancaiani, A., A.F. Anfossi, A. Barbara, S. Benedetto, B. Blasi, V. Carletti, *et al*. (2015) Evaluating scientific research in Italy: The 2004-10 research evaluation exercise. *Research Evaluation*, 24: 242-255.

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

Anderson, D.L., W. Smart, W., & J. Tressler (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.

Auranen, O. & M. Nieminen (2010) University research funding and publication performance – An international comparison. *Research Policy* 39: 822-834.

ANVUR (2013) *Rapporto finale. Valutazione della qualità della ricerca 2004-2010 (VQR 2004-2010)*. Roma.

Baccini, A. & G. De Nicolao (2016) Do they agree? Bibliometric evaluation versus informed peer review in Italian research assessment exercise. *Scientometrics,* 108: 1651-1671.

Barker, K. (2007) The UK Research Assessment Exercise: The Evolution of a National Research Evaluation System. *Research Evaluation*, 16: 3–12.

Boyle, G. (2008) Pay Peanuts and Get Monkeys? Evidence from Academia. *The B.E. Journal of Economic Analysis and Policy*, 8: 21.

Broadbent, A. (2010) The UK Research Assessment Exercise: Performance Measurement and Resource Allocation. *Australian Accounting Review*, 52: 14–23.

Buckle, R.A. & J. Creedy (2019a) The evolution of research quality in New Zealand universities as measured by the performance-based research fund process. *New Zealand Economic Papers*, 53: 144-165.

Buckle, R.A. & J. Creedy (2019b) An evaluation of metrics used by the performance-based research fund process in New Zealand. *New Zealand Economic Papers*, 53: 270-287.

Buckle, R.A. & J. Creedy (2020) The 'Disciplinary Effect' of the Performance-Based Research Fund Process in New Zealand. *New Zealand Economic Papers*, 54: 107-126.

Buckle, R.A., J. Creedy & N. Gemmell (2020) Is external research income associated with convergence or divergence of research quality across universities and disciplines?

Evidence from the PBRF process in New Zealand. *Applied Economics*, 52: 3919-3932.

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

Checchi, D., M. Malgarini & S. Sarlo (2019) Do performance-based research funding systems affect research production and output? *Higher Education Quarterly*, 73: 45-69.

Checchi, D., I. Mazzotta & M. Malgarini (2020) Convergence or polarisation? The impact of research assessment exercises in the Italian case. *Scientometrics*, 124: 1439-1455.

Dale, T. & S. Goldfinch (2005) Article citation rates and productivity of Australasian political science units 1995-2002. *Australian Journal of Political Science*, 40: 425-434.

de Boer H., B. Jongbloed, P. Benneworth, L. Cremonini, R. Kolster, A. Kottmann, K. Lemmens Krug & H. Vossenstey (2015) *Performance-Based Funding and Performance Agreements in Fourteen Higher Education Systems.* Report for the Ministry of Education, Culture and Science, Centre for Higher Education Policy Studies, C15HdB014, *Universiteit Twente, Enschede*.

Ehrenberg, R., H. Kasper & D. Rees (1991) Faculty turnover at American colleges and universities: Analyses of AAUP data. *Economics of Education Review*, 10: 99–11.

Franzoni, C., G. Scellato & P. Stephan (2011) Changing incentives to publish. *Science*, 333: 702-703.

Gemmell, N., P. Nolan & G. Scobie (2017) Estimating Quality-Adjusted Productivity in Tertiary Education: Methods and Evidence for New Zealand. *Victoria University of Wellington Working Papers in Public Finance*, No. WP17/2017.

Geuna, A. & B.R. Martin (2003) University research evaluation and funding: An international comparison. *Minerva*, 41: 277-304.

Hare, P.G. (2003) The United Kingdom's Research Assessment Exercise: impact on institutions, departments, individuals. *Higher Education Management and Policy*, 15: 43-62.

Hicks, D. (2012) Performance-based university research funding schemes, *Research Policy*, 41: 251-261.

Jung, J. (2012) Faculty Research Productivity in Hong Kong across Academic Discipline. *Studies in Higher Education*, 2: 1-13.

Kolarz, P., E. Arnold, F. Dijkstal, K. Nielsen & K. Farla (2019) *International landscape study of research and innovation systems.* Brighton. Technopolis.

Martin, B.R. (2011) The Research Excellence Framework and the impact agenda: are we creating a Frankenstein monster? *Research Evaluation*, 20: 247-254.

Moore, W.J., R. J. Newman, P.J. Sloane & J.D. Steely, (2002), Productivity Effects of Research Assessment Exercises. *University of Aberdeen Centre for European Labour Market Research Discussion Paper*, 2002-02.

OECD (2010) *Performance-Based Funding for Public Research in Tertiary Education Institutions: Workshop Proceedings*. Paris: OECD.

Payne, A.A. & J. Roberts (2010) Government oversight of public universities: are centralized performance schemes related to increased quantity or quality? *The Review of Economics and Statistics*, 92: 207-212.

Sabharwal, M. (2013) Comparing Research Productivity across Disciplines and Career Stages. *Journal of Comparative Policy Analysis: Research and Practice*, 15:141–163.

Schumpeter, J.A. (1942) *Capitalism, Socialism and Democracy.* London: Routledge.

Shin, J.C., & W.K. Cummings (2010) Multilevel Analysis of Academic Publishing across Disciplines: Research Preference, Collaboration, and Time on Research. *Scientometrics*, 85: 581–594.

Shorrocks, A.F. (2013) Decomposition procedures for distributional analysis: A unified framework based on the Shapley value. *Journal of Economic Inequality*, 11: 100-126.

Stone, R. (1965) A model of the educational system. *Minerva*, 3:172-186.

Stone, R. (1966) Input-output and demographic accounting: a tool for educational planning. *Minerva*, 4: 365-380.

Smart, W. (2009) The impact of the Performance-Based Research Fund on the research productivity of New Zealand universities. *Social Policy Journal of New Zealand*, 34: 136-151.

The Tertiary Education Commission (2019) *Improving research quality: The results of the PBRF 2018 Quality Evaluation.* Wellington: Tertiary Education Commission.

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

The Tertiary Education Commission (2004) *Performance-Based Research Fund Evaluating Research Excellence: the 2003 assessment.* Wellington: Tertiary Education Commission.

van Besselar, P., U. Heyman & U. Sandstrom (2017) Perverse effects of output-based research funding? Butler's Australian case revisited. *Journal of Informetrics*, 11: 905–918.

Wanner, R.A., L. S. Lewis & D. L. Gregorio (1981) Research productivity in academia: A comparative study of the sciences, social sciences and humanities. *Sociology of Education*, 54: 238–253

Wang, J. & D. Hicks (2013) Detecting structural change in university research systems: A case study of British research policy. *Research Evaluation*, 22: 258-268.

Wilsdon J., L. Allen, E. Belfiore, P. Campbell, S. Curry, S. Hill, …, & B. Johnson (2015) *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. London: Sage.

Woerlert, P., & L. McKenzie (2018) Follow the money? How Australian universities replicate national performance-based funding mechanisms. *Research Evaluation*, 27: 184-195.

Xu, Y.J. (2008). Faculty turnover: Discipline-specific attention is warranted. *Research in Higher Education*, 49: 40–61.

About the Authors

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

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

Ashley Ball is an Honours Student in Economics at Wellington School of Business and Government, Victoria University of Wellington, New Zealand.
Email: ballashl@myvuw.ac.nz