The Performance Based Research Fund in NZ: Taking Stock and Looking Forward

Robert A. Buckle and John Creedy

WORKING PAPER 02/2022
January 2022

Working Papers in Public Finance

VICTORIA UNIVERSITY OF WELLINGTON TE HERENGA WAKA

Chair in Public Finance

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

The Performance Based Research Fund in NZ: Taking Stock and Looking Forward

By

Robert Buckle and John Creedy

Victoria University of Wellington, New Zealand

Abstract

This paper draws on earlier research by the authors to review changes in research quality in New Zealand universities since the introduction of the Performance Based Research Fund (PBRF) in 2003. The changes are related closely to the incentives created by the scheme, and are associated with the nature of the considerable staff turnover that has taken place over the 15-year period during which it has operated. The precise funding formulae used, relating to the research funds attached to different discipline groups and quality categories, involve political judgements and are not considered here. However, a review of the changed nature of universities and the details of the evaluation process suggest that substantial simplifications could usefully be made while maintaining incentives that are at the heart of any PBRF.

JEL classifications: I2; I22; I28; L32.

Acknowledgements: We are grateful to the New Zealand Tertiary Education Commission (TEC) for providing the anonymised longitudinal data used here. The data are not publicly available and were provided by the TEC following a confidentiality agreement. We should like to thank Norman Gemmell for comments on an earlier draft.

1. Introduction

New Zealand universities have operated within a Performance-Based Research Fund (PBRF) since 2003. It is one component of a 'dual-support' system of research funding. The performance-based component is allocated to a university, based on a research assessment process, and the university has discretion over the use of those funds. Prior to the PBRF, such funding was provided on a bulk funding process based on student numbers. The second component is contestable research funding sought from independent bodies and earmarked for specific research projects. The NZ system is just one of many performance-based schemes, implemented in a range of countries during the last four decades: for an international review, see Hicks (2012).

Three reasons tend to influence the adoption of a performance-based system for public funding of research institutions. These are: the need to ensure stable and reliable long-term financing; the need for accountability; and a desire 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. An early and influential example is the UK's Research Assessment Exercise introduced in the 1980s, subsequently renamed the Research Excellence Framework; on its early development, see Martin (2011).

The main features of the NZ PBRF have remained constant over the three full-scale reviews carried out in 2003, 2012 and 2018. Following the last evaluation exercise, an official review panel was established, resulting in PBRF Review Panel (2020). The final recommendations largely involved increasing funding for specified ethnic groups (by increasing subject area weights and funding weights) and broadening the definition of research. It was proposed to appoint a Sector Reference Group that, 'demonstrates a strong commitment to Maori-Crown partnership and comprises a diverse membership'. The Panel also recommended that the Tertiary Education Commission should 'discontinue reporting the Average Quality Score metrics' for discipline areas and universities. The Review Panel thus concentrated on a narrow range of political and funding allocation aspects. It is suggested that this was a missed opportunity to take a deeper look at the system and its contribution to the stated aims.

The aim of the present paper is therefore to conduct a broader stocktake. It is based on a series of detailed analyses by the authors, which make use of the extensive data generated by the

PBRF over a 15-year period; see Buckle and Creedy (2018, 2019a, 2019b, 2020), Buckle, Creedy and Gemmell (2020, 2021) and Buckle, Creedy and Ball (2021). The stocktake involves analysing the nature of the changes that have taken place in NZ universities in relation to the initially-stated objectives of the PBRF. In considering whether the design of the PBRF is consistent with its objectives, it is necessary to analyse the nature of the incentives, facing university managers and individual researchers, created by the process. Of particular importance is the nature of the metric used to measure research quantity. Furthermore, it is necessary to examine whether the changes that have taken place have produced an environment in which it is appropriate to continue to use the same metrics, or whether changes are warranted.

The background is provided in Section 2, which outlines the origins of the NZ PBRF. The stated objectives of the system are discussed in Section 3. Section 4 discusses the measurement of research quality, highlighting a number of unusual and undesirable features of the unique system. Section 5 reviews the effects of the PBRF on research quality in NZ. Here special attention is given to the fundamental processes which contribute to improvement: these involve staff turnover and quality improvements by researchers who remain in the same university over time. Section 6 considers how the global rankings of NZ universities, which necessarily use different metrics, have changed since the introduction of the PBRF. Changes in postgraduate training are briefly discussed in Section 7. Having considered the changes since the introduction of the PBRF, Section 8 looks forward with a number of suggested modifications. Brief conclusions are in Section 9.

2. Origins of the NZ Performance Based Research Fund

The statutory origin of the NZ PBRF appears to be a 1998 White Paper which made a series of recommendations concerning tertiary education. These included: more and better information for students, providers and Government; improved accountability for research funding, governance, quality assurance and audit; and changes in research funding. Rather than allocating research funding from Vote Education, via subsidies per equivalent full-time student (EFTS), the White Paper recommended allocating funding via a contestable pool for advanced research. Tertiary research funding by Government was included in the bulk funding of tertiary education institutions on the basis of EFTS, adjusted by weighting for different course costs. This bulk funding was intended to cover capital and operating costs, as well as tuition and research. In allocating funding, there was little attention to accountability, capacity building, and governance.

After the 1999 general election, the Labour Government established a Tertiary Education Advisory Commission (TEAC) to consider strategic issues for the tertiary sector, including the principles which should underpin research funding for the sector. In a series of reports, the TEAC recommended the introduction of a performance-based research fund, and the establishment of funds to support research centres and networks of excellence: see New Zealand Ministry of Education (2018a, pp. 15-17). The TEAC recommendations were motivated by research performance, inadequate research funding and the need for incentives. It was also concerned about lack of accountability for the use of public funds for research, lack of concentrated research funding, substantial differences in the allocation of research funding among disciplines, and inadequate funding for research training. The TEAC also recommended the establishment of the Tertiary Education Commission (TEC), which administers the PBRF.

The TEAC reviewed tertiary funding schemes in several other countries, and proposed a research funding model based on three elements: a Quality Evaluation (QE) of all individual eligible academic staff (weighted at 50 per cent of the fund), a measurement of external research income (ERI) (weighted at 25 per cent), and a measurement of postgraduate research degree completions (RDC) (weighted at 25 per cent). It recommended that the quality rating of academics should be based on a mix of performance-indicators (such as bibliometric measures) and peer-review. It also recommended that quality ratings be determined by institutional self-assessment, subject to five-yearly external audits (of a random sample of about 10 per cent of staff), conducted by independent, multi-disciplinary assessment panels.

A Working Group (PBRFWG) was established in 2002 to advise the Transition TEC and Ministry of Education on the detailed design and implementation of a PBRF. This group recommended a two-step process to evaluate the quality of researchers, but this was not adopted. The PBRFWG also recommended a broad assessment of individual performance, based on research output, peer esteem and contributions to the research environment. A further suggestion was that a higher weight (60 per cent) of the research funding be based on the individual academic staff research assessment component (QE) and a lower weight (15 per cent) for the external research income (ERI) component. For a comprehensive account of the process and issues considered during the development of New Zealand's PBRF, see Boston (2006).

The system introduced in 2002 comprised the following features that have remained largely unchanged throughout three full assessment rounds in 2003, 2012 and 2018, and a partial round in 2006. As proposed by the TEAC, three measures are used to allocate Government funding

to support research at universities and other Tertiary Education Organisations (TEOs). For 2003 and 2012 the weighting for each component was 60 per cent for the QE component, 25 per cent for the RDC component, and 15 per cent for the ERI component. For 2018 the weights were 55 and 20 per cent for QE and ERI respectively. The funding allocation via the QE component is based on an assessment of the research quality of each eligible staff member, based on a submitted Research Portfolio. This assessment process, described further in Section 4 below, involves each researcher being assigned to a Quality Category (QC) by a complex peer-review process undertaken by a panel of experts in each subject area. These individual QCs are used to allocate funding, and to compute average performance scores for subject areas and institutions.

3. Objectives of the New Zealand PBRF

There have been several statements of the purpose of the introduction of a PBRF to the New Zealand tertiary education sector. However, the following quotation, which has been reiterated in numerous Ministry of Education and TEC documents, would seem to capture the overall aim:

> 'The purpose of the ... [PBRF] ... is to ensure that excellent research in the tertiary education sector is encouraged and rewarded. This means assessing the research performance of tertiary education organisations (TEOs) and then funding them based on their performance'. (Tertiary Education Commission, 2019)

Elsewhere, the New Zealand Ministry of Education (2018b) has elaborated in more detail what 'excellent research' should comprise and promote. In a paper prepared for the 2020 Review Panel, the Ministry explained that:

> 'The primary objectives of the PBRF are to: (i) Increase the quality of basic and applied research at New Zealand's degree-granting tertiary education organisations (TEOs); (ii) Support world-leading teaching and learning at degree and postgraduate levels; (iii) Assist New Zealand's TEOs to maintain and lift their competitive rankings relative to their international peers; and (iv) Provide robust public information to stakeholders about research performance within and across TEOs.'

It also states that in doing so, the PBRF will also: (v) support the development of postgraduate student researchers and new and emerging researchers, (vi) support research activities that provide economic, social, cultural, and environmental benefits to New Zealand, including the advancement of Mātauranga Māori, and (vii) support technology and knowledge transfer to

New Zealand businesses, iwi and communities. These objectives are to be pursued consistent with a set of guiding principles that include statements about comprehensiveness, respect for academic traditions, consistency, continuity, differentiation, credibility, efficiency, transparency, complementarity, and cultural inclusiveness: see New Zealand Ministry of Education (2018b, pp. 1-2).

Following the Final Report, TEC is in the process of appointing a Sector Reference Group for the 2025 evaluation, charged with designing and implementing the changes recommended by the Panel. These include promoting, 'a more holistic approach to research excellence', and adding a new objective, 'to ensure a flourishing and inclusive system for developing and sustaining research excellence'. They also recommend three new principles to 'better reflect the distinctive partnership between the Crown and Māori'.

A difficulty with the broadly stated objectives is that, despite the complex numerical research quality metric discussed in the following section, no guidance is actually provided of what value of the metric would be regarded as a good achievement for NZ. Given the uniqueness of the measure adopted, consistent international comparisons are difficult to make. The statement of objectives is extremely vague, making it difficult to provide an assessment of whether the costs of the scheme are outweighed by its benefits. There is an additional difficulty arising from the fact that, as discussed in Section 2, there are many more objectives ('targets') than the three measured outcomes on which research funding is based ('instruments').

4. The measurement of research quality by the PBRF

This section describes the metric used to measure research quality, which is needed to understand the incentives created by the PBRF. Only when these incentives are articulated is it possible to assess whether, and to what extent, the changes in NZ universities since 2003 can actually be attributed to the PBRF.

There is a vast literature concerned with the measurement of research quantity and quality. In introducing the PBRF, a highly complex and unique assessment method was introduced, although no rationale has ever been provided for the idiosyncratic, and indeed somewhat eccentric, features of the quality evaluation method used. This is unfortunate as it may reasonably be expected that, in introducing such a substantial and complex policy evaluation process, an explanation would be provided of how the measurement relates to the policy objectives.

When considering the nature of a university career, it is clear that a distinct life-cycle pattern exists. Entrants to the profession, following post-graduate training, devote much energy to establishing early research projects, and developing a portfolio of publications in refereed journals. This is combined with attendance at conferences, and presentations to colleagues in other universities, which, in combination with publications, help to establish a reputation and network of contacts. In the early stages of a career it is difficult to obtain substantial research grants, and it can take some time for research quality to be recognised and to generate indicators such as citations and invitations to participate in wider professional activities.

As academics become established they take on a greater diversity of roles, often involving the leadership of teams of researchers, and the supervision of graduate students. In addition they are in a position to participate in the wider activities of the profession, such as becoming editors of journals or taking on responsible positions in professional organisations. Their work may also attract the attention of those outside universities, such as professionals working in the public sector. In some cases, substantial administrative roles may be taken on, necessarily involving research trade-offs.

While a typical kind of career progression can be described in this way, it is necessary to recognise substantial differences among individuals in their relative skills, aptitudes and personal characteristics. It cannot be expected that even the best researchers can, or would wish, to cover at a high level *all* the possible types of activity that exist within a university career.

However, this heterogeneity, which most people would surely consider a desirable feature of a university, does not appear to have been recognised in the PBRF metrics used to measure research quality, which involve a fixed weighting to different types of activity. The constraints facing young academics are recognised only in the designation of the additional letters, NE, for ‘new and emerging’ researchers, after the quality category assigned by the peer review process. Researchers are expected to achieve excellence across a wide range of indicators, and there is little room for trade-offs, which are themselves a necessary feature of individual (and indeed public) decision making. This unreasonable requirement makes it particularly difficult to achieve the highest quality category of the PBRF, as illustrated below, even for highly productive researchers who have strong national and international reputations in their chosen fields of specialisation.

The quality evaluation process is described briefly as follows. The information submitted by each researcher is referred to as an 'Evidence Portfolio' (EP). This is evaluated by subject panels to derive the 'Quality Category' (QC).

There is inevitably the important wider question of how quality is actually judged by the peer-review panel members. This can involve decisions about the extent to which judges are expected to read research papers (or view other types of research output) and the possible use of external indicators, such as journal rankings and impact scores. Of relevance here, for example, is the use in a wide range of business and social science disciplines, of the *ABCD Journal Quality List* compiled by the Australian Business Deans Council, and which places journals in four categories. The list generates considerable controversy, particularly regarding the incentive to target research towards higher-ranked journal, to reduce the range and innovation of research, with resulting disincentives to work on 'local' policy problems; see Willmott (2011) and Chatterjee *et al.* (2020). Furthermore, the list does not correlate well with other indicators; see Hirschberg and Lye (2020). Such difficulties alone suggest that there are dangers in attempting to use a fine categorisation of research quality, with numerical scores, as in the NZ PBRF. These broader issues cannot be discussed here, where the emphasis is on evaluating the *effects* of the PBRF via the incentives created by the form of metrics used.

Each portfolio is assigned a score from 0 to 7 for each of three components: 'research output'; 'peer esteem'; and 'contribution to research environment'. These three scores, s_j, are given weights, q_j, of 0.70, 0.15 and 0.15. The Total Weighted Score, S, for an individual is obtained by multiplying the weighted sum of the s_j values by 100. Hence:

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

The scores take only integer values and the maximum individual score is 700. It may be expected that, after assigning a score for each individual, an arithmetic mean score could be obtained for a university or discipline group, with each individual value weighted by the full time equivalent (FTE) status. However, instead of taking such an obvious direct route, the PBRF system converts each researcher's value of S into a letter grade, as follows.

Four different quality categories are distinguished. These are as follows: R for scores 0 to 199; C for scores between 200 and 399; B for scores from 400 to 599; and A for scores from 600 to 700. A numerical score, G, is then assigned to each letter grade: 10 for an A; 6 for a B; 2 for a

C; and 0 for R. In 2012 the weight for C(NE) was the same as for C. However, in 2018 the weight for C(NE) was increased to 4. The weight for R(NE) remained zero over both periods.

The Average Quality Score (*AQS*) for a university or discipline group is then defined as the FTE-weighted average of the relevant *G*s. No explanation was provided as to why all the detailed information regarding scores, *S*, running from 0 to 700, was compressed into four values of *G,* or why the particular values for the *G*s (which relate to the funding amounts) or *q*s (which relate to quality measurement) were selected.

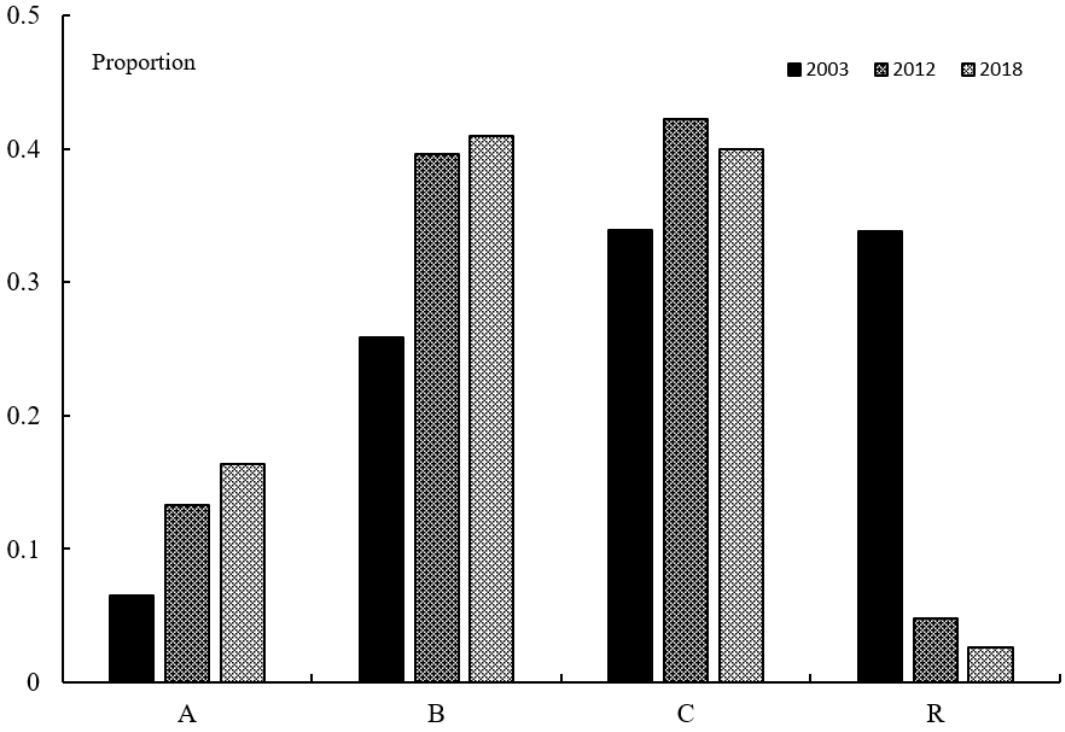
For example, suppose two universities have exactly the same distribution of quality categories; that is, the proportions of As, Bs, and Cs are the same in each university. However, one university has all its researchers at the top end of the scale for each category, while the other university has all researchers at the bottom end of each relevant scale. They have identical *AQS*s, and funding per researcher, and yet there are actually considerable differences between their assessed research qualities. This naturally raises the question of what was the point in producing *S* values for each individual if only the QCs matter. And why produce *AQS*s, given that they are irrelevant for the funding rules?

There are further oddities, which do not appear to be widely recognised. Although *S* can range from 0 to 700, there are nevertheless only 512 values that it can take, although some of these can be achieved in several different ways. The distribution has rather strange properties, as shown by Buckle and Creedy (2019b). For example, there are 101 integer values of *S* between 600 and 700, with 200 values between 400 and 599. This may suggest that there are twice as many ways to get a B category compared with an A. Yet, if all individuals were to have an equal probability of obtaining any value from 0 to 7 for each of the three components, only 6.6 per cent of individuals would achieve an A, whereas 35.0 per cent would achieve a B: the B grade is over five times more common than the A grade, not twice as common. Furthermore, in this equal-probability case, the chances of obtaining B, C or R grades are not equal, despite being associated with apparently equal ranges of *S*. It can be shown that 35.7 per cent would be assigned a C grade, while the remaining 22.7 per cent would obtain an R grade. The peculiarities of the system are explored further in Buckle and Creedy (2019b), who also examine the iterations taken by the peer review panels, involving large adjustments to initial scores, in arriving at the final QC categories for individuals.

In 2018 the number of components was reduced to two: ‘research output’ and ‘research contribution’, with the latter having a weight of 0.3. The difficulty of achieving an A score can

be illustrated in this case as follows. Consider the score needed for the research output component, in order to achieve a quality category of A, for a given score assigned to research contribution. If the latter is 3 or less, it is impossible to obtain A. If research contribution is given either 4 or 5, it is necessary to obtain the maximum score of 7 for research output, in order to achieve a weighted average giving an A quality category. With a 6 or 7, then obviously an A requires at least 6 for research output. The ability to achieve an A-quality category is therefore considerably reduced by the limited ability, arising from the use of integer scores, to trade-off high research output for even a slightly lower research contribution.

Figure 1. Distribution of portfolios by QC: 2003, 2012 and 2018



Source: Authors' calculations using TEC data

From the point of view of funding, the *AQS* (and thus the number attached to each quality category) is completely redundant, since what matters are the numbers of FTEs falling in each of the quality categories. These are shown, for all universities and discipline groups combined, in Figure 1 for each of the PBRF rounds. This diagram clearly displays the major changes that have taken place, involving the large reduction in Rs, along with the increase in the proportion of As and Bs, with an initially increasing proportion of Cs, followed by a reduction. This is consistent with the incentives created by the metrics, as discussed further below, since Cs are less attractive as the *AQS* rises above the weight (of 2) attached to them. These are consistent

with the incentives discussed above. Furthermore, after the initial changes from 2003 to 2012, subsequent improvements were clearly much harder to achieve. There are substantial differences between discipline groups and universities, as shown in Buckle and Creedy (2019b, 2020).

There is an additional perverse feature of the PBRF metrics. Many other metrics of research performance have been devised over many years. Despite their differences, they all have the same basic feature, in that they give rise to a positively skewed distribution of performance. It is even possible to provide a statistical rationale for this property, as first shown by Shockley (1957). Where research performance requires a number of attributes (such as having original ideas, perseverance, communication skills, a willingness to respond to criticism, and so on), a positively skewed distribution arises if these different attributes operate multiplicatively. The multiplicative nature is expected because, for example, original ideas combined with little perseverance is not likely to produce high performance. However, the distribution of scores, *S*, is nothing like a positively skewed distribution; see Buckle and Creedy (2019b).

The publication of the *AQS* measures, for universities and subject areas, nevertheless relates to objective (iv) mentioned above, to provide public information to stakeholders about research performance within and across TEOs. The TEC publishes results for each TEO, discipline and subject after each PBRF assessment round. Nevertheless, the 2020 PBRF Review Panel expressed concerns that the *AQS* measure is 'not necessarily compatible with the objective of providing robust public information' (PBRF Review Panel 2020, p. 5) and have recommended that the reporting of the *AQS*s should be discontinued (Recommendation 33, 91).

One reason given for this recommendation is that it was considered that the various *AQS* metrics are difficult to interpret. Yet the Panel did not, it seems, recognise any need to modify the metrics used, or provide clear information which would make interpretation easier. If the measures are considered to be incompatible with the objective of providing robust information, the response should surely be to modify and clarify the measures, rather than simply to stop their publication. It is argued below that much clearer information would instead be provided by reporting QC distributions for each discipline group within each university.

5. The effects on the research quality of New Zealand universities

This section is concerned with evaluating the effects of the PBRF on the measured research quality, as measured by the *AQS* of NZ universities. This clearly relates to objective number (i) listed above, although in stating the objective, no guidance was provided as to what the TEC

regarded as indicating a high-performing university or department. In discussing the possible effects of the PBRF, one important consideration cannot be overlooked. The fact that performance measures corresponding to the metrics do not exist before the introduction of the PBRF raises a challenge in attributing any changes over the various 'rounds' to the PBRF itself. For example, other incentives exist, involving, say, the reputations of individual researchers and universities, and the desire to obtain competitive research grants. In addition, the rewards offered by the PBRF may, for some discipline groups and universities, be small in relation to other funding opportunities.

Nevertheless, with careful consideration of the precise nature of the new incentives introduced by the PBRF, and examination of detailed data collected over the 15 years of its experience in NZ, it is possible to make a strong argument that the changes within universities and discipline groups have closely reflected those incentives. Buckle and Creedy (2019a) listed a set of hypotheses regarding the changes within universities that would be expected to arise from the incentives created by the PBRF. Statistical analyses showed strong support for those hypotheses.

This inference is made possible by the fact that the PBRF involves an evaluation of each researcher. The resulting anonymised data allow for the detailed changes in the structure of universities to be examined. Without such detailed information, the connection between incentives and changes would not be so clear.

It is important to recognise the fundamental point that an improvement in the *AQS* of any research group comes about via two primary mechanisms.

- The first is turnover of staff: this involves (on average) the exit of lower-quality researchers and the entry of higher-quality researchers.
- The second mechanism is the improvement, or quality transformation, of existing staff.

A possible third way of improving the *AQS* of a university, though not a discipline, is to change the discipline composition by reducing the size of lower-quality discipline groups, and expanding that of higher-quality groups. However, there are substantial constraints on the ability of NZ universities to make large changes in their discipline composition. For instance, all programmes must be approved by the Committee on University Academic Programmes, administered by the NZ Vice Chancellors' Committee. Furthermore, the TEC determines which programmes are to be funded, and can refuse funding for programmes deemed to be inappropriate. In addition, teaching commitments prevent substantial variations in staff-student

ratios from taking place. Using a decomposition of quality changes, Buckle, Creedy and Ball (2021) showed that this provided a negligible contribution to universities' improved *AQS*s. Nevertheless, some changes, particularly the contraction of education, have been consistent with this mechanism.

There are limitations to the ability to transform individuals from low- to high-quality researchers, through the use, for example, of internal grants and other mechanisms to encourage research. Universities also face constraints regarding the 'management out' of poor-performing members of staff, and there are market constraints regarding the ability to recruit higher-quality researchers. These constraints include the quite rigid salary structures in universities, and the higher rewards provided by non-academic, or 'outside', opportunities for individuals in particular discipline groups, which vary among disciplines.

Given the two main mechanisms mentioned above, of turnover and quality transformations, the appropriate technique for examining changes is the social accounting framework. This 'maps' the precise nature of the movements, or transitions, of individuals between different categories over time. Examples of such transitions are shown in Table 1, for all universities combined for the periods 2003 to 2012, and 2012 to 2018.

Table 1. Matrix of researcher 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.
Source: These proportions are taken from a larger table in Buckle, Creedy and Ball (2021, p. 214).

Using this approach for each university, it has been found that the pattern of exits, entrants and transformations has varied significantly among universities: they all differ from each other in important ways; for universities see Buckle and Creedy (2019a) and for discipline groups see Buckle and Creedy (2020).

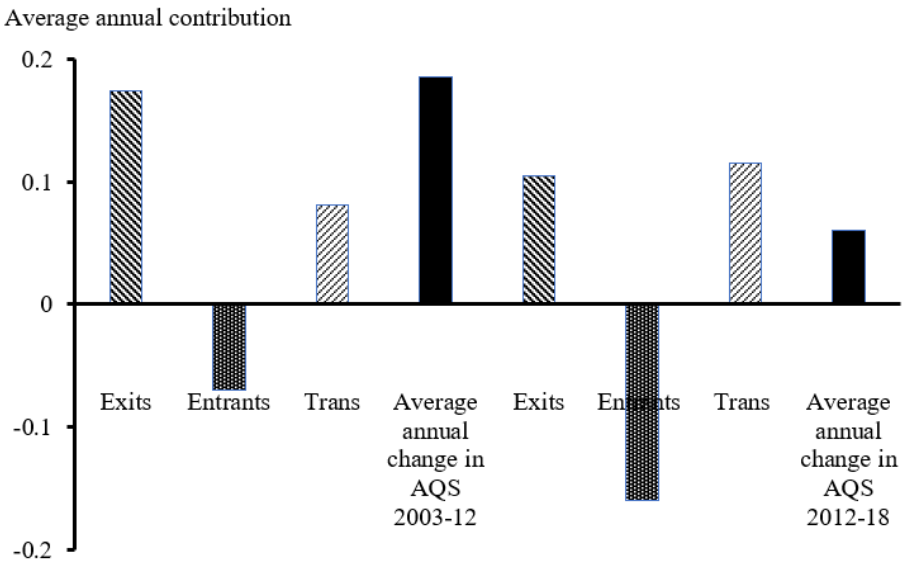
However, all universities share a common over-riding characteristic, in that their transitions are consistent with the incentives created by the metrics used by the PBRF. In particular, a university with an *AQS* above 2 must recruit A and B researchers if entrants are to raise its *AQS*, whereas a university with *AQS* < 2 can raise its score by recruiting C-type researchers. Furthermore, lower-*AQS* universities have an initially higher proportion of R-quality staff, and can more easily raise their score by encouraging the exit of substantial numbers of Rs. Therefore, it is expected that universities – given their different starting points – will display different patterns of change, while conforming to a common theme.

The evidence shows that the recruitment of higher-quality researchers has been greater among universities with relatively higher initial *AQS*s, which are clearly more attractive to A and B-type researchers. Nevertheless, such entries have been relatively low, reflecting both the scarcity of such researchers and, in some fields, the inability to offer competitive salaries. Compared with quality transformations and exits, this has meant that the contribution of staff entrants to *AQS* changes has been consistently negative over the more recent PBRF experience (in the sense that entrants, at the date of the subsequent PBRF, are on average of lower quality than incumbents).

The contributions of exits, entrants and quality transformations to average annual *AQS* changes are shown in Figure 2 for all universities combined. Similar decompositions can be obtained for separate universities and discipline groups; see Buckle, Creedy and Gemmell (2021).

The negative effect of entrants means that the average research quality of entrants was lower than the end-of-period *AQS* of those researchers who remained in the system. The *AQS* of entrants can nevertheless be above the *AQS* at the beginning of the period. In the second period, the higher initial *AQS* (in 2012) made it more difficult to recruit entrants above the average quality. Unless entrants have an *AQS* greater than or equal to the final *AQS* of incumbents, they cause the *AQS* to fall.

Figure 2. Contributions of exits, entrants and quality transformations to average annual changes in AQS for all universities



Source: Buckle, Creedy and Gemmell (2021).
Notes: The values for "Average annual change in *AQS* …" are equal to the sum of the values for exits, entrants and transformations shown to the left.

In looking at the factors leading to *AQS* improvements, there are two outstanding features of staff turnover in NZ universities over the PBRF period, as follows.

- The first characteristic is the very high overall *amount* of staff turnover: this is not surprising in view of the fundamental role of turnover in generating aggregate quality change, and the difficulty of achieving quality transformations, particularly at the higher levels.
- Secondly, the single most important characteristic of turnover has been the exit of a large number of R-quality researchers. This high exit rate has meant that exits have contributed the largest component of *AQS* improvement for all universities.

The nature of these transitions revealed by the social accounting framework, and the fact that the differences observed across universities and discipline groups are consistent with the incentives created by the PBRF, combined with the overall market scarcity of very high-quality researchers who can be attracted to NZ universities, has two further important and closely associated implications.

- First, the introduction of the PBRF has generated large-scale changes in NZ universities that are not sustainable over a long period. If the initial changes are projected to

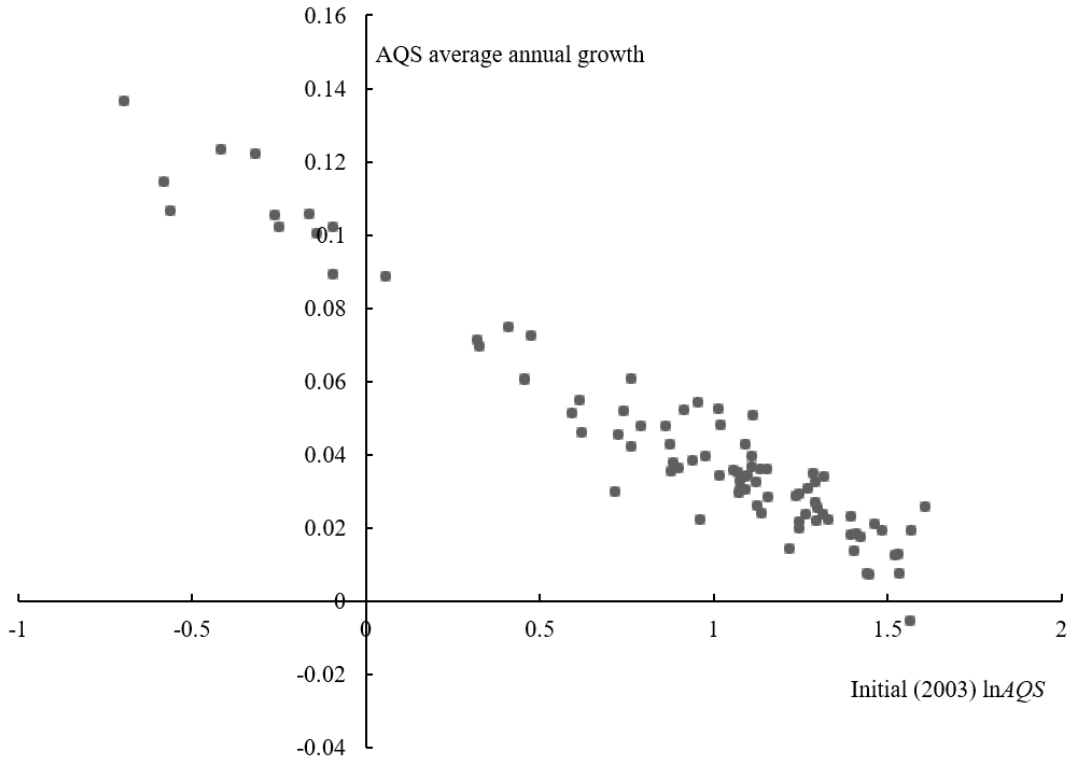
continue over a much longer period, they result in unrealistic numbers of academics, as demonstrated by Buckle and Creedy (2019a). Furthermore, because the responses to the incentives depend on the initial *AQS* level, those initial responses (such as the recruitment of relatively large numbers of C-type researchers) eventually – as *AQS* levels improve – contribute negatively to further change. Importantly, the rate of increase in *AQS*s has slowed down over the more recent PBRF round covering the period from 2012-2018 compared with earlier PBRF rounds covering the period from 2003-2012. This feature, a diminishing effect on average research quality, alone raises the question of whether it is worthwhile continuing with the PBRF in its present highly complex and cumbersome form.

- A second, and closely related, implication is that the metrics used by the PBRF generate an inbuilt tendency for quality convergence among universities and discipline groups. This convergence process involves two characteristics.
    - First, there is a tendency for lower-quality universities to experience higher *AQS* growth rates, compared with higher-quality universities. Hence, it can be said that the changes have involved a type of systematic 'catch-up' process. In addition, the extent of the catch-up process, while substantial, has been restrained somewhat in recent years by the difficulty in sustaining improvements, mentioned earlier. This means that those universities with relatively higher *AQS* growth between two earlier periods, experienced – among those with similar *AQS* levels at the start of the second period – slightly lower subsequent growth.
    - Secondly, the dispersion of average research quality across universities and discipline groups has declined. This is not a necessary consequence of the first convergence characteristic, but results from the dominance of that first characteristic over any extraneous inequality-increasing features. Hence, in terms of their *AQS*s, NZ universities have become more alike as a result of the PBRF. This contrasts with a situation in which research excellence becomes concentrated in a small number of places.

The extent of convergence over the period of the PBRF, 2003 to 2018, is illustrated in Figure 3. This shows the annual average *AQS* growth rate, plotted against the logarithm of the initial (2003) *AQS*, for universities and discipline groups. There are 87 observations in all: with 8 universities (all disciplines); 9 disciplines (all universities); plus $8\times9 - 2 = 70$

for the disciplines within universities, allowing for the absence of Law and Education in Lincoln university; the compositions of the discipline groups are listed in Buckle and Creedy (2020). These observations are clearly arranged around a downward-sloping straight line, showing the tendency for lower-*AQS* universities and subjects to experience, on average, higher subsequent *AQS* growth. The diagram also reveals that a similar convergence process applies to all universities and discipline groups. These properties are examined in detail in Buckle, Creedy and Gemmell (2020, 2021).

Figure 3. Initial AQS and average annual AQS growth, 2003 to 2018



Source: Authors calculations using TEC data

In summary, it is possible to trace a substantial general increase in the *AQS*s of NZ universities, along with a strong convergence process and associated 'decreasing returns', and to understand, via the social accounting framework, the precise dynamics – in terms of the nature of staff turnover and quality transformations – that have contributed to this growth. Furthermore, the stimulus to these changes can be understood in terms of the incentives created by the PBRF, and the various constraints on change, particularly imposed by the nature of the

academic labour market in NZ. This type of exercise is a crucial part of any evaluation of the policy.

Nevertheless, it is not possible to provide a measure of the extent to which this could be judged to be a policy success or failure. It is not even possible to make like-for-like international comparisons, given the unique nature of the PBRF. For example, one may reasonably ask whether the improvement (in terms of its own metrics) has been sufficient to warrant the administrative and compliance costs, which themselves are difficult to measure. This inability to evaluate such a large change in the conditions facing universities is of course shared with the many policy initiatives. Policy makers seldom, if ever, provide a clear indication of the desired outcomes, other than general expressions that cannot be used as clear tests of performance.

6. The global ranking of New Zealand Universities

Direct *AQS* comparisons with overseas universities are impossible, despite the stated aim of improving the global rankings of NZ universities. It is therefore worth considering the rank changes obtained using other metrics, and to compare them with changes measured by the PBRF. One ranking system is produced by Quacquarelli Symonds (QS): http://www.iu.qs.com/university-rankings/world-university-rankings/. This includes 'academic reputation' based on a global survey, and citations per academic, along with the ratio of faculty to students, employer reputation, the proportion of international students and the proportion of international faculty. 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.

Details of the QS world rankings are given in Table 2. In 2012 the QS ranking was, in order from first to last: Auckland (AU), Otago (OU), Canterbury (CU), Victoria University of Wellington (UW), Massey (MU), Waikato (WU), Lincoln (LU), and Auckland University of Technology (AUT). These ranks are significantly different from the PBRF-based ranks. However, virtually all the difference arises from the position of VUW. The lower QS rank for VUW could perhaps arise from other factors which outweigh any gains in research-related quality. Furthermore, 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.

Table 2: PBRF and QS World University Rankings of NZ Universities

University	2012		2019				2022		
	PBRF	QS	PBRF	QS			QS		
	AQS	Global rank	*AQS*	Global rank	AR	C/f	Global rank	AR	C/f
VUW	5.38	237	5.34	221	43.6	47.6	236	46.7	37.5
AU	4.89	83	5.38	85	82.7	52.3	85	83.8	55.1
OU	4.85	133	4.86	125	48.6	50.9	194	50.6	49.5
CU	4.61	221	5.05	231	35.6	52.9	258	39.2	41.1
WU	4.21	374	4.95	274	18.3	71.7	373	19.9	44.5
MU	4.11	308	4.56	332	27.3	24.9	284	29.9	23.7
LU	3.51	na	4.54	317	7.8	39.7	372	8.3	30.2
AUT	3.15	451-500	3.73	464	17.5	10.6	451	20.3	15.6

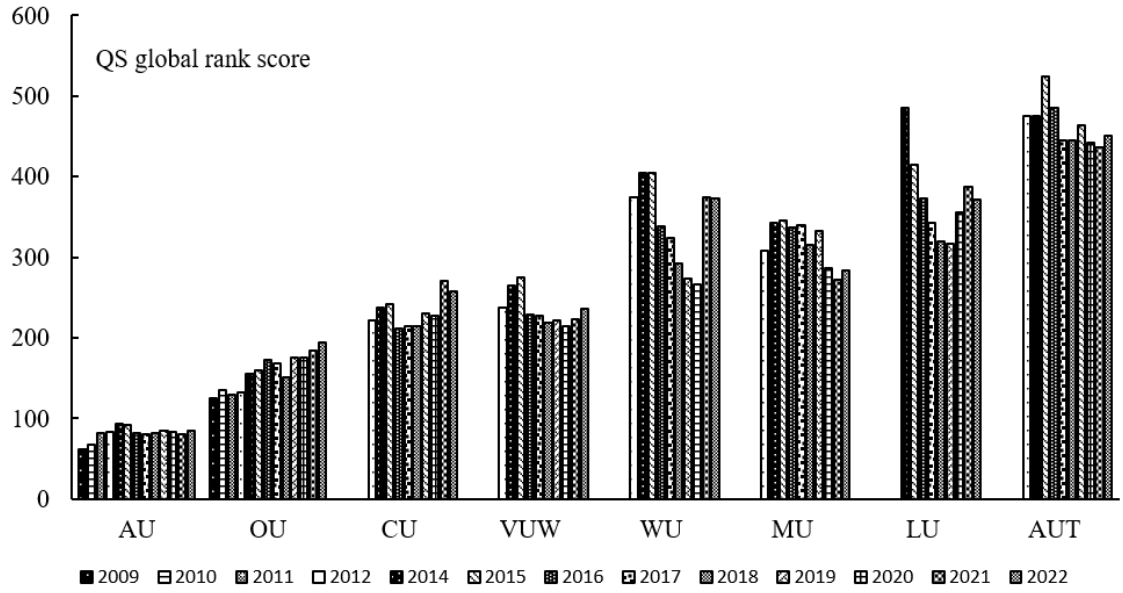
Sources: PBRF scores are from Buckle, *et al*., (2021); QS data are from Quacquarelli Symonds (2021).
Notes: Universities are listed in order of PBRF ranking based on *AQS* for the university.

Figure 4 shows the movements in the QS global ranking of New Zealand universities from 2009 to 2022. Initially, QS published only the top 200 ranked universities. Only three New Zealand universities were captured in the QS from 2009 to 2012. These were Auckland (AU), Otago (OU) and Canterbury (CU), for 2009 and 2010 only. From 2012 onward (except 2013) QS have published the rankings for all New Zealand universities. Figure 4 shows a variable pattern in the trend for the QS global ranking scores. For the highest QS ranked university, AU, the score deteriorated slightly from 2009 to 2015, and since then it has remained fairly stable. The second highest QS ranked university, OU, has seen a steady decline in its ranking over the 14-year period. The QS ranking for CU has also declined slightly. For the other five universities there has been, in general, a moderate improvement in the QS rank scores.

The QS rankings are determined from scores for six metrics: Academic Reputation (40%), Citations per faculty (20%), Faculty to student ratio (20%), Employer reputation (10%) International faculty ratio (5%), and International student ratio (5%). Table 2 includes the 2012 and 2018 *AQS*s for each NZ university and the QS global ranking for 2012, 2019 and 2022 (there is a QS ranking for all years from 2012 to 2022, except 2013). It also shows scores for each of the two components of interest that contribute to the QS ranking: Academic Reputation (AR) and Citations per faculty (C/f). In both cases the QS scores seem likely to be more heavily influenced by experience and the age of a researcher (other things equal) than the RO component of an individual researcher's score in the PBRF, though they may have a closer correspondence with the other components (CRE and PE). The AR score is based on a survey of over 130,000 individuals in higher education, regarding teaching as well as research quality.

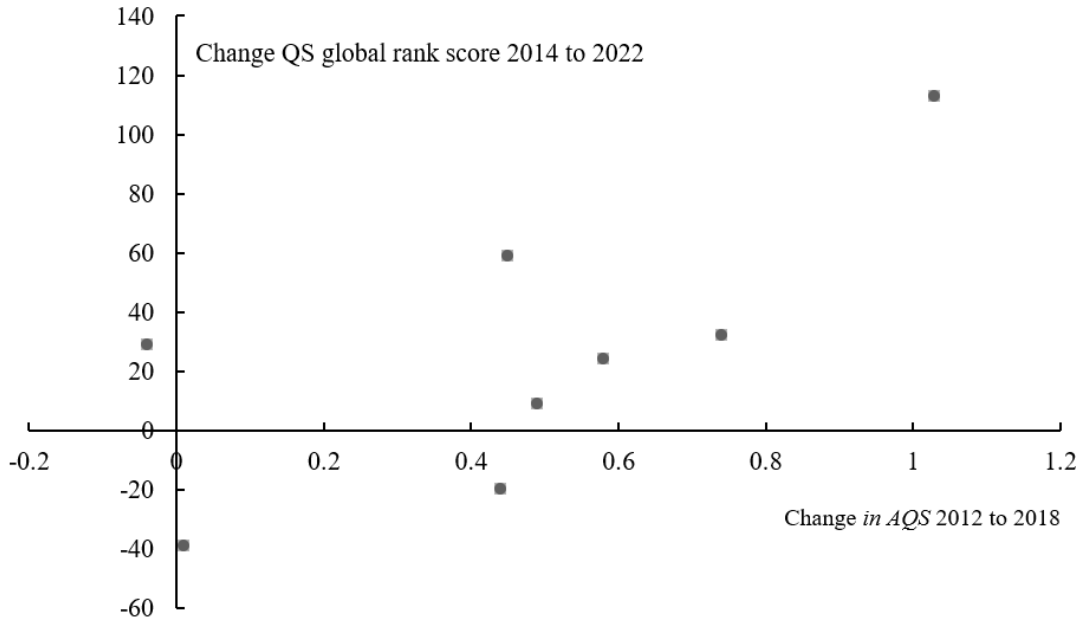
The C/f score is based on the total number of citations received by all papers produced by an institution across a 5-year period of faculty members of the institution.

Figure 4: QS global ranking of New Zealand Universities, 2009-2022



Source: Quacquarelli Symonds (2021).
Notes: The rank scores for AU are the mid-points of a range from 2012 to 2018; WU and LU are mid-point of a range for 2014 and 2015.

Figure 5: Changes in QS global ranking of New Zealand Universities and changes in PBRF research quality scores.



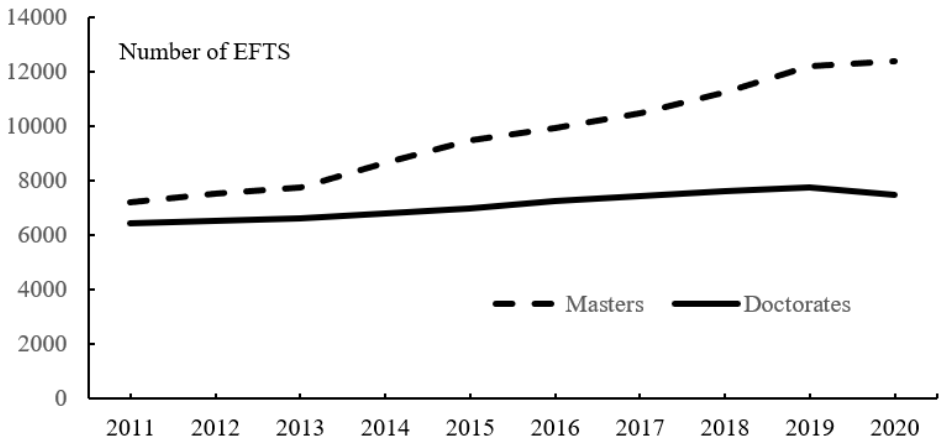
Source: Authors' calculations

Given the relatively heavy weight given to research indicators, it is interesting to explore the relationship between the PBRF performance of NZ universities and changes in their QS global rankings, notwithstanding the limited comparability of the two measures. Figure 5 plots the change in the PBRF-based *AQS* between 2012 and 2018 and the change in QS global ranking for each New Zealand university, suggesting that the changes are positively related. This therefore provides some *prima facie* evidence that the PBRF has contributed to a general improvement in global rankings.

7. Teaching and learning at postgraduate levels.

In addition to changes in research quality, and incentives to increase external research income, the PBRF created incentives to increase the number of research degree completions (RDC). Figures 6 plots the total number of Masters and Doctorate effective full-time students (EFTS) enrolled in NZ universities from 2011 to 2020.

Figure 6: Total EFTS of Masters and Doctorates at NZ universities



Source: New Zealand Ministry of Education: Domestic and international equivalent full-time student units (EFTS) by sub-sector, provider and qualification type/NZQF level 2011-2020

The growth in the number of Doctorate enrolments is moderate. As a proportion of total university EFTSs it has increased from 4.8 per cent to 5.4 per cent from 2011 to 2020. The Masters EFTSs show stronger growth, but this includes both professional as well as research degree enrolments.

On the relationship between the research achievement of NZ academics and the labour market outcomes of students, Khurana *et al*. (2021) use PBRF data on the research activity of individual university academics and the departments of majoring subjects. Their results show there is no systematic difference between more and less research-active departments in terms

of students' labour market outcomes. Khurana *et al.* summarise an extensive body of research on the link between research productivity and teaching quality, though only a limited number of studies have investigated the link between teaching quality and subsequent student outcomes.

Evidence on the overall effect of research on teaching is mixed, with some studies finding a positive link, others a negative one, and some suggesting that the two scholarly activities are largely independent. Evidence also suggests that the relationship between teaching and research differs across disciplines, the level of study, workload allocations, faculty-student ratios and institutional settings.

8. Looking forward: suggested modifications to the PBRF

Having examined the nature of changes taking place since the PBRF began, and the new incentives created by it, this section proposes a number of changes to the process. An important consideration is the finding that there have been 'decreasing returns', reflected in a recent slowing down of the rate of research quality improvement. Following initial large improvements in *AQS*s over early years of the PBRF, which largely resulted from the exits of large numbers of research inactive, R-category, staff, subsequent growth rates were much lower. Associated with those decreasing returns, there has also been a catching-up process in which there is now a lower degree of variation in research quality among universities.

Hence the 'research quality profile' of New Zealand universities is very different from the situation when the scheme began in 2003. In addition, the real value of government research funding has declined over the period, with a growing reliance on international student revenue, which has recently declined substantially as a result of the border restrictions imposed during the Covid-19 pandemic. One response by university managers has been to offer voluntary redundancy and retirement, a consequence of which has been an earlier-than-otherwise loss of more senior researchers. The implications for research output and quality are uncertain.

There are also strong constraints imposed on universities aiming to improve their research quality. These involve: the scarcity of very top-quality researchers; (the associated) pay differentials between university careers and 'outside' options; the need for a balance of younger/experienced staff in a dynamic system; the need to maintain teaching commitments across discipline areas; and limited access to competitive research funding sources for some disciplines.

In considering the future of PBRF evaluations and funding formulae, it is necessary to recognise, as stressed above, that incentives play a crucial role. The changes taking place in NZ universities have been found, through a detailed analysis of the resulting demographic transitions, to be consistent with the incentives created by the PBRF. Hence, it is not suggested here that the PBRF should be abandoned. Indeed, it is important to maintain clear incentives for individuals, departmental heads, and senior university managers. However necessary incentives could be maintained with a much-simplified process, which is thereby also more transparent.

The changes suggested below would also reduce transaction and compliance costs for the TEC, university administrators, and academics. Hazledine and Kurniawan (2005) suggest a variety of costs are likely to be incurred including: administrative costs incurred by the administering agencies (TEC and Ministry of Education), compliance costs incurred by the universities, rent seeking costs incurred by individuals, departments, faculties and universities, and the costs of improving research performance that would be incurred by universities. The changes suggested below are likely to influence the first two categories of costs. These categories of costs, and the costs of rent seeking behaviour, fall within the deadweight loss type of costs. It is of course difficult to obtain reliable information on these costs. Hazledine and Kurniawan report Web of Science (2004) detailed estimates of the total of costs incurred by the administering agencies and compliance costs of universities for the first PBRF round in 2003. The Web of Science estimate is less than 2 per cent of the total funds to be allocated, which they report was comparable with Hong Kong and the UK.

Potentially the more important additional costs impacted by the PBRF is the category ‘Costs of improving research performance’. This category includes the change in average salary per academic resulting from hiring better quality researchers, and the cost of allocating more of a university’s resources toward supporting research. These costs are more reflective of the process of enhancing the average research productivity of researchers and universities. But estimating the marginal effect on these costs of introducing the PBRF is clearly difficult. It is not obvious how the changes suggested below would impact these costs, other than by providing more accurate information about research performance.

The suggested changes are listed below, with a brief explanation and rationale. These suggestions relate to the research evaluation process, rather than the separate issue of funding weights or amounts, for which different types of consideration apply. The process of allocating funding is essentially a value-laden process which inevitably involves much lobbying

behaviour. But such decisions can be separated from the question of quality measurement. Hence no attempt is made here to suggest funding weights, or amounts, for different subject groups.

Before considering changes, it is worth stressing two basic aspects which should remain.

1. ***Retain individual evaluations***. This is necessary to maintain incentives, and to provide information for university managers as well as the TEC. Importantly, it is has been shown that, for evaluation purposes, information about individual flows in a social accounting framework is crucial for identifying the precise sources of universities' research quality changes.
2. ***Retain peer evaluation***. Evidence has shown that quantitative measures, such as those provided by bibliometric measures of publications and citations, do not correlate well with the broader-ranging quality metrics used in peer evaluations. Of course, peer reviews are not without their own problems, but have long been a fundamental part of universities regarding student and staff evaluations. However, the present peer evaluation process is extremely cumbersome and time consuming and, as shown above, much of the process is not necessary.

The suggested modifications are listed as follows.

1. ***Assess all university non-administrative staff who do not have teaching-only contracts***. In a research evaluation exercise, it makes no sense to exclude any staff who are expected to engage in research, irrespective of their FTE status or the nature of their contracts. The PBRF appears to have stimulated the use of teaching-only academics by some universities, and it would be useful to record separately the number of such staff.
2. ***Abandon the use, in Evidence Portfolios, of 'self assessment'***. The present system encourages researchers to engage in self-promotion. This is not necessary in a peer review process.
3. ***Eliminate the use of metrics s, q, and S (defined in Section 4) which are used to assign individuals to quality categories, QC.*** These metrics give the false impression that precise cardinal measures of research quality, allowing comparisons among individuals and subject areas, can be made. Importantly, they have important limitations, in terms of their numerical properties, and no rationale for them has ever been given. As discussed above, the current system starts by assigning – for each portfolio – numerical values to several 'types of research output' and then forms a weighted sum, which has strange properties.

The resulting number is then converted into one of the four discrete QCs (although that conversion process involves numerous adjustments over several stages). Eliminating these metrics also involves abandoning the use of separate types of research categories, which in the earlier rounds were 'research output', 'peer esteem', and 'contributions to research environment'. The peer review process could instead proceed directly as follows.

4. ***Directly assign staff to quality categories***. Something like the present categories could continue to be used: these are R, C, B and A, along with C(NE) if it is decided to provide different funding 'rewards' for C and C(NE). The category, R(NE), is redundant given the zero-funding allocated to R's. It is important to provide clear guidelines about what is considered to be research output, and then use peer review to assign individuals directly to the different categories (which essentially refer to 'no substantive output', low, medium, and high quality). In order to achieve consistency, and to provide information to judges and researchers, examples of portfolios which correspond to the different categories need to be provided for each subject area. In setting out such examples, it is also necessary to avoid confusing outputs with inputs. For example, research grants are inputs (though success in obtaining grants may depend on reputation and the perceived local importance of the research topic). An important concomitant of this suggestion is the following change.
5. ***Introduce peer review Panels for each discipline, rather than broad discipline groups***. Portfolios that are considered to represent 'high' or 'medium' research quality, for example, necessarily vary considerably among discipline groups. This is because of, among other things, the vastly different publishing practices, and attitudes to conference presentations, books and journal publications. Research quality comparisons among disciplines involve huge difficulties, and crucially it is not necessary to presume that they can be made. The present system also gives the false impression that cardinal comparisons are possible.
6. ***Replace the AQS with a reported distribution of individuals over the research quality categories***. At present, after the assignment of staff to QCs, a score is attached to each category. For research quality aggregation purposes, there is no rationale for the particular scores used, although they currently relate, if loosely, to funding amounts per FTE staff member. (In addition, the Final Report of the Review Panel controversially proposed awarding different funding amounts depending on the declared ethnicity of the researcher). In the final stage, an FTE-weighted average score, the *AQS*, within each discipline group, and within each university, is calculated. There is no need for the production of such an average score, since funding anyway depends on the number of FTEs in each separate category, and the discipline area. All that matters for funding purposes, and the need to

provide information, is the number of FTEs in each category. Given the distribution of researchers across the various categories, it is of course possible for anyone to impose their own subjective weights and to compute some type of average quality measure for disciplines and universities. But for the purposes of the PBRF, there is no need to pretend that weights have anything to do with objective quality measurement, rather than funding allocations. Indeed, in something as subjective as quality measurement, it seems desirable to be as free as possible from subjective weighting schemes.

9. Conclusions

New Zealand academics and university managers, since 2003, have operated with a Performance Based Research Fund system which allocates money to universities according to a complex funding formula that is itself related to an even more complex peer evaluation exercise. Following the latest evaluation round in 2018, the process has been subject to an official Review, carried out under the aegis of the Ministry of Education. Unfortunately, that review concerned itself only with proposals to broaden the definition of research, and to increase funding for specified subject areas and researchers, depending on their declared ethnicity.

It is argued here that, in view of the substantial changes that have taken place over the PBRF period, an analysis of the detailed structure of the evaluation process is warranted. This paper has therefore carried out a stocktake of the review process, based on an earlier series by the authors of detailed analyses of data generated by the PBRF, revealing the dynamics of the NZ university system. Not least among the changes is the substantial amount of staff turnover, which can be directly related to the incentives created by the PBRF. This turnover is essentially because changes in the overall research quality of a university can arise only through exits, entrants and quality transformations of incumbent researchers (given the constraints on changes in its discipline composition).

Although there have been substantial variations among disciplines and universities in the precise nature of the changes, identified using a social accounting framework, it has nevertheless been established that those differences are consistent with the incentives. Put simply, for example, relatively high-performing universities can increase their quality measure only by recruiting high-quality researchers, whereas lower-quality universities can improve their overall measure by recruiting medium-quality researchers. And the initial composition of non-research active staff differs among universities and discipline areas. NZ universities have

gone through a catch-up process in which the between-university variation in measured research quality has fallen in response to a strong systematic tendency for lower-quality universities to improve, on average, faster than initially higher-quality universities.

Given the large changes since 2003, along with the more recent substantial shock to universities as a result of policy responses to Covid-19, it is worth considering whether the PBRF evaluation process is 'fit for purpose', and whether changes could usefully be made to the peer review procedures used to arrive at quality measures. It is suggested here that the arcane method used to allocate researchers to Quality Categories should be abandoned, and replaced by a simple direct ordinal judgement, with no pretence that a (highly idiosyncratic) cardinal scale can be applied to suggest that, for example, an A researcher has five times the research quality of a C researcher. The process also needs to ensure that judgements are made by 'peers' within the discipline of the researcher being evaluated, not by a panel made up of a variety of discipline groups with widely differing research cultures.

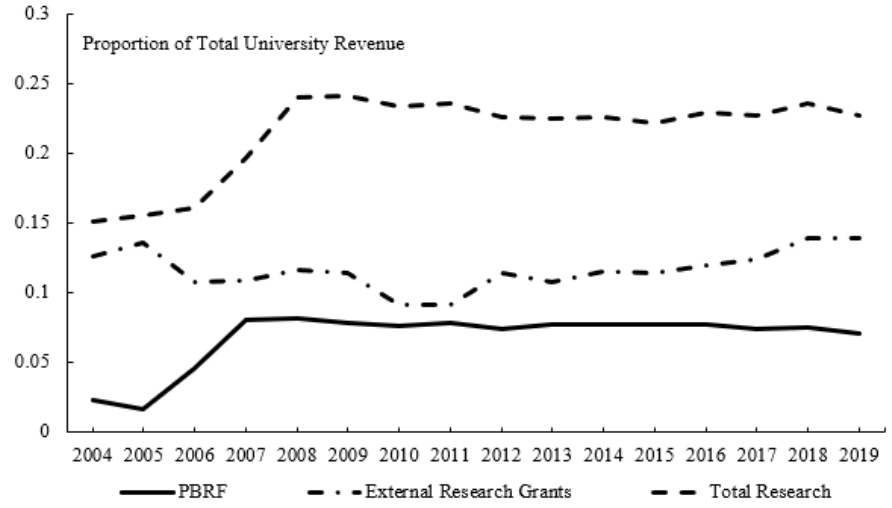
The approach to quality-measurement can be separated from the funding formulae used. Given a distribution of researchers over the different quality categories, an independent judge may impose relative weights to the different categories in order to produce some kind of aggregate quality index (while recognising that different judges may not agree). But when deciding on the funding to be provided, a quite different set of considerations are relevant, involving, for example, costs, scarcity of skills, value judgements, and political considerations, which are likely to vary among discipline groups. The present approach conflates these two aspects by using spurious quantitative measures of research quality, which are then directly linked – with various adjustments – to the funding formulae.

A great deal has been learned about New Zealand universities, their substantial responses to the introduction of explicit research incentives, and the application of a unique 'measuring rod' used to assess research quality. The university environment is very different from that in existence when the PBRF was introduced in 2003. It would be remarkable if a policy evaluation found no fault with the process, or failed to recognise the nature of the substantial staff turnover stimulated by it. Changes along the lines suggested here are warranted in order to produce a simpler and more transparent peer review process that is likely to involve lower compliance and evaluation costs, while maintaining desired incentives.

Appendix A: Financial incentives from the PBRF

Figure 7 shows total revenue received by the eight New Zealand universities in each year from for PBRF performance (the aggregate of revenue received for QE, ERI and RDC); from External Grants and Total Research revenue, each as a proportion of total university revenue. After establishing the PBRF system and after the transition to the full PBRF revenue allocation process, total revenue received from the PBRF reached 8.1 per cent of total university revenue and has subsequently settled at just over 7 percent. Note that this is largely the consequence of switching from the process operating prior to 2004 allocating revenue, including research revenue, via a bulk funding process based on FTES numbers. Hence, total identifiable research revenue increased to about 24 per cent of total revenue in 2008 and has settled at around 23 per cent. The share of External Research Grant revenue increased from 2011 to 2019 to be about 14 per cent of total revenue, but this is no higher than in 2004.

Figure 7: Components of University Research Revenue as a Proportion of Total Revenue

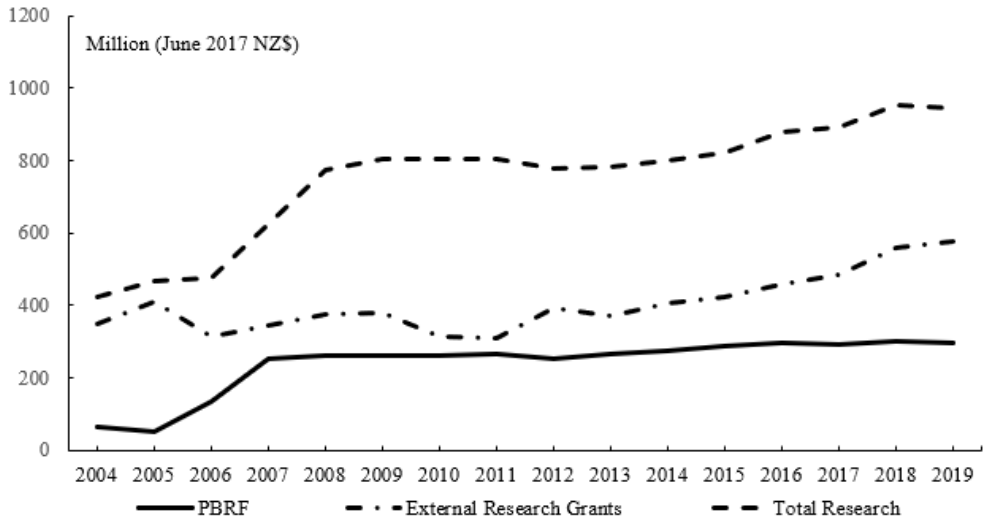


Source: TEC: Financial performance | Tertiary Education Commission (tec.govt.nz)

Perhaps more helpful is the growth in the real value of components of research revenue, which is shown in Figure 8. This suggests that after the initial allocation of funds via PBRF the real value of PBRF revenue and Total Research revenue remained largely unchanged until after 2013. Thereafter, the growth in Total Research revenue has been the result of growth in the real value of External Research Grants while the real value of PBRF revenue has remained unchanged. Hence, perhaps the growth in the quality of NZ university researchers coupled with the incentives created by the ERI component of the PBRF has induced growth in revenue from

External Research Grants. However, the growth in Total Research revenue has been slower than the growth of Total University revenue. The real value of Total University revenue grew by about 29 per cent between 2008 and 2019, whereas Total Research revenue grew by about 22 per cent (of which External Research Grants revenue grew by 54 per cent but PBRF revenue grew by only about 12.5 per cent).

Figure 8: The Real Value of Components of University Research Revenue



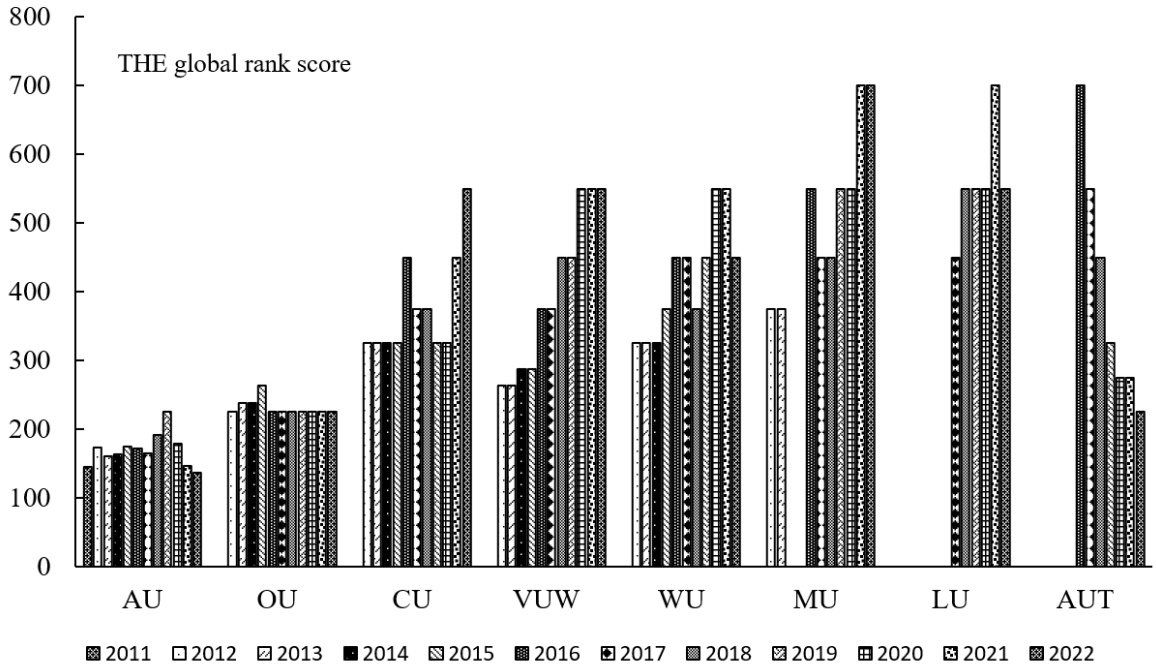
Source: Derived from TEC: Financial performance | Tertiary Education Commission (tec.govt.nz) and Statistics New Zealand data.

After the initial allocation change from a bulk-funding system based on student numbers to a funding system based on performance, the real value of total PBRF revenue received by universities has remained largely unchanged. This is despite a growth in funded PBRF portfolios between 2003 and 2018, and a faster growth in PBRF funded researchers and higher quality researchers. However, the total PBRF revenue includes revenue from all three components (QE, RDC and ERI) and not simply the QE component.

Appendix B: New Zealand Universities and THE World Rankings

Information about world rankings is available from the Times Higher Education (THE) ranks, which are based on a weighting of performance indicators grouped into five areas. Weights are: 30% for Teaching; 30% for Research; 30% for Citations; 7.5% for International outlook; and 2.5% for Industry income. Figure 9 shows the THE global rank for NZ universities, not all of which were included until 2017. The pattern is markedly different from the QS global ranking scores in Figure 4. In the THE system, only AUT has shown a systematic improvement; Auckland (AU) and Otago (OU) have broadly maintained their ranking while the others have declined. This pattern contrasts with changes in PBRF scores and ranks since 2003.

Figure 9: The THE global ranking of New Zealand Universities, 2011-2022



Source: Data are from World University Rankings 2022 | Times Higher Education (THE) (2021).

The pattern of change for the 'research component', which makes up 30 per cent of the overall score for a university, is quite different from that for the global rank scores. The research component scores are shown in Figure 10. The pattern is, in general, one of steady, and in some cases continuous, improvement. However, the THE Research score for Auckland University (AU) relative to the scores for the other universities, particularly for VUW, OU an CU, is an outlier compared with what would be expected, based on the PBRF *AQS* scores.

Figure 10: The THE 'Research' scores for New Zealand Universities, 2011-2022

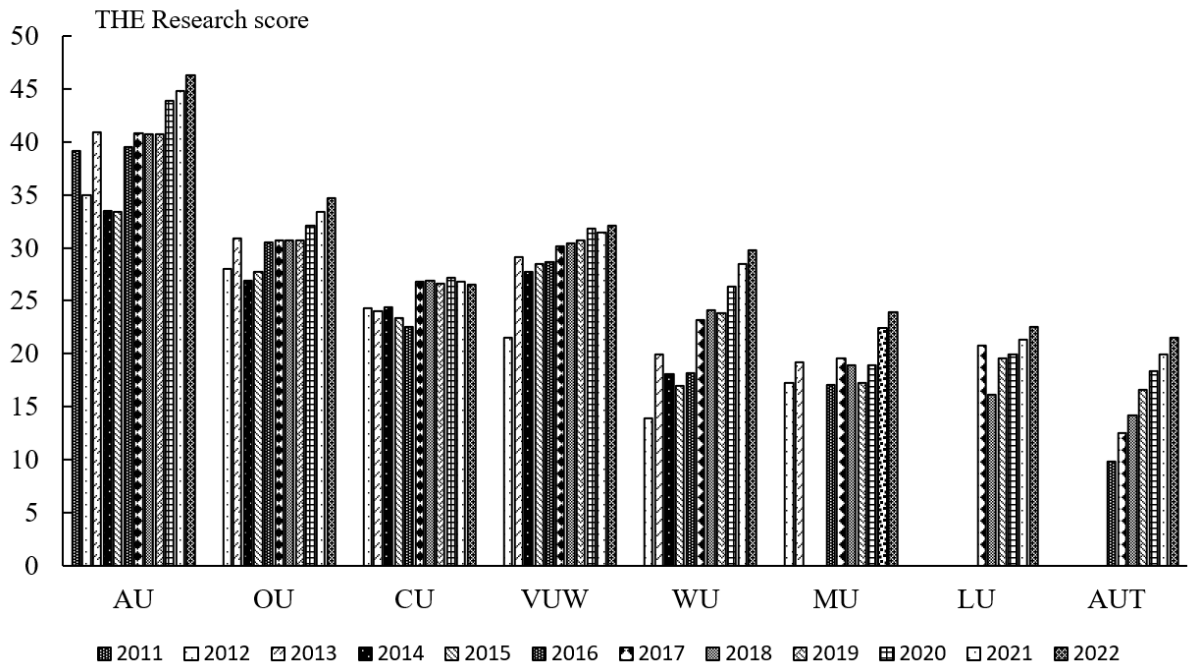
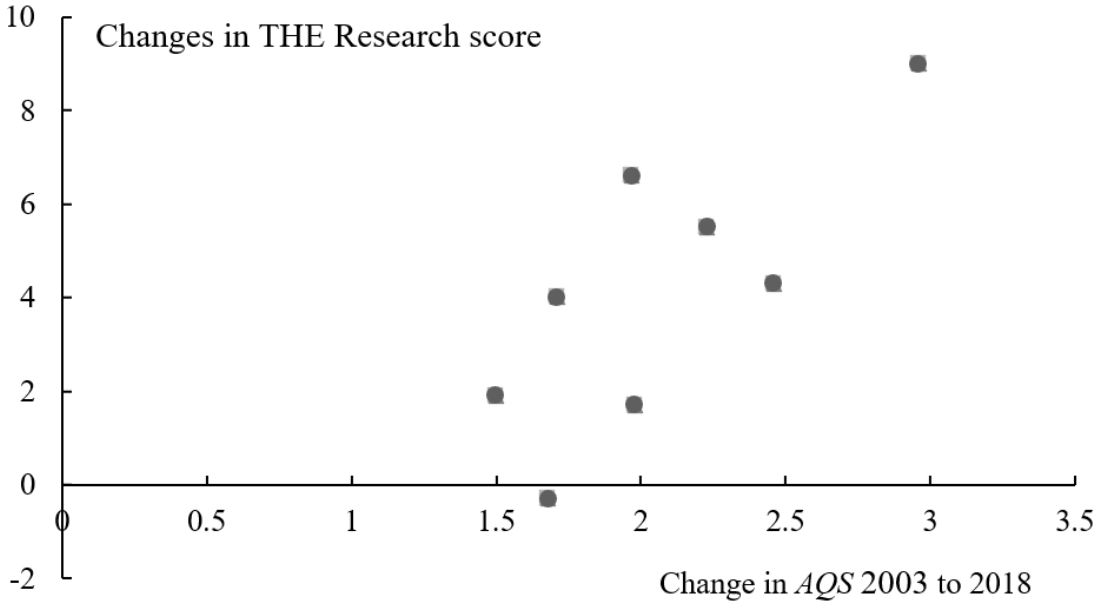


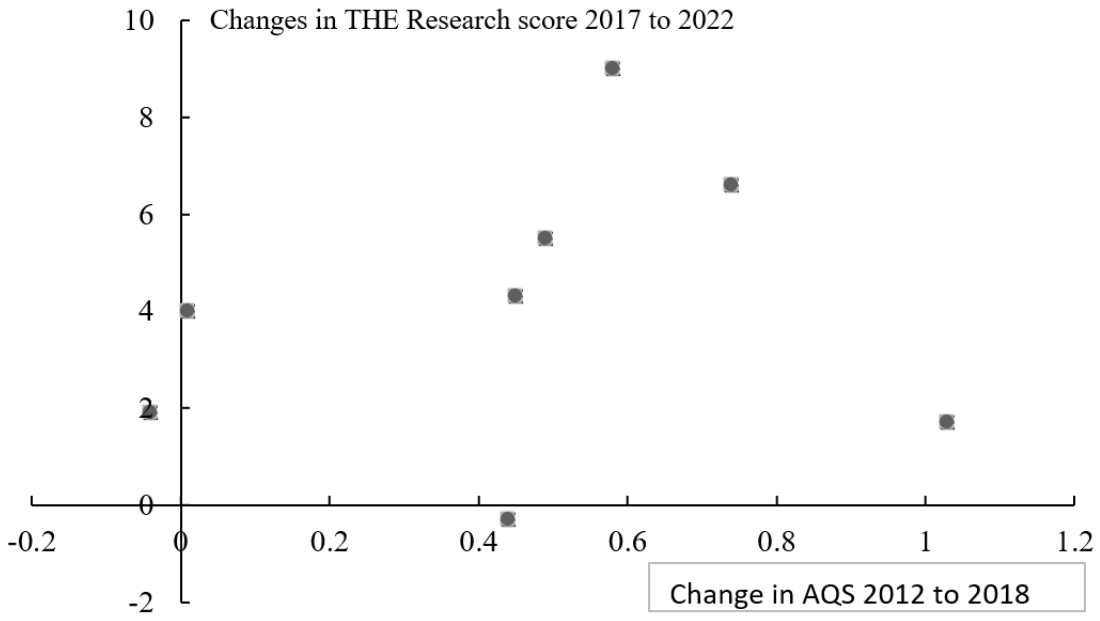
Figure 11: Improvement in THE Research score of New Zealand Universities (2017 to 2022) and changes in PBRF research quality scores (2003 to 2018).



Source: Authors' calculations

Figure 11 shows the scatter-plot of the change in the PBRF-based *AQS* between 2003 and 2018 and the (subsequent) change in THE Research score for each New Zealand university between 2017 and 2022. Figure 12 shows the scatter-plot of the change in the PBRF-based *AQS* between 2012 and 2018 and the (subsequent) change in THE Research score for each New Zealand university between 2017 and 2022. Similar patterns are evident in the relationships between changes in Citation scores between 2017 and 2022 and changes in *AQS*s for each university, although the relationship is not as strong.

Figure 12: Improvement in THE Research score of New Zealand Universities (2017 to 2022) and changes in PBRF research quality scores (2012 to 2018).



Source: Authors' calculations

References

Buckle, R.A. and Creedy, J. (2018) The impact on research quality of performance-based funding: the case of New Zealand's PBRF scheme. *Agenda*, 25: 25-48.

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

Buckle, R.A., Creedy, J. & Gemmell, N. (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.

Buckle, R.A., Creedy, J. & Gemmell, N. (2021) Sources of convergence and divergence in university research quality: Evidence from the Performance-Based Research Funding System in New Zealand. *Victoria University of Wellington Working Papers in Public Finance*, WP05/2021.

Buckle, R.A., Creedy, J. & Ball, A. (2021) Fifteen years of a PBRF in New Zealand: incentives and outcomes. *Australian Economic Review*, 54: 208-230.

Boston, J. (2006) Rationale for the Performance-Based Research Fund and its Evaluation. In Bakker, L. Boston, J. Campbell, L. and Smyth, R. (Eds), *Evaluating the Performance-Based Research Fund: Framing the Debate,* Wellington: Institute of Policy Studies.

Chatterjee, B., Cordery, C.J., Loo, I.D. & Letiche, H. (2020) The spectacle of research assessment systems: insights from New Zealand and the United Kingdom. *Accounting, Auditing and Accountability Journal*, 33: 1219-1246.

Hazledine, T. & Kurniawan, C. (2005) Efficiency of New Zealand universities and the impact of the performance based research fund. *New Zealand Economic Papers*, 39: 153-179.

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

Hirschberg, J.G. and Lye, J.N. (2020) Grading journals in economics: the ABCS of the *ABDC*. *Journal of Economic Surveys*, 34: 876-921.

Khurana, S., Gunby, P. and Menclova, A. K. (2021) The Relationship between Research Achievement of Academics and Student Labour Market Outcomes. Paper presented to the NZ Association of Economists' Annual Conference, Wellington. Retrieved from: 623971_8d8db8380044440e9cd6c73228c8667d.pdf (filesusr.com)

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

New Zealand Ministry of Education (2018a) *A History and Overview of the PBRF*. Wellington: Ministry of Education. Retrieved from: https://www.education.govt.nz/assets/Documents/Further-education/Policies-and-strategies/Performance-based-research-fund/PBRFHistoryAndOverview.pdf

New Zealand Ministry of Education (2018b) Terms of Reference Issue Paper 1: Revisiting the objectives of the PBRF. Retrieved from: Terms-of-Reference-Issue-Paper-1-Revisiting-PBRF-Objectives.pdf (education.govt.nz)

PBRF Review Panel (2020) *Towards the Tertiary Research Excellence Evaluation (TREE): The Report of the PBRF Review Panel*. Wellington: Ministry of Education.

Quacquarelli Symonds (2021) *QS World University Rankings*. Retrieved from: QS World University Rankings 2022: Top Global Universities | Top Universities

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

WEB Research (2004) *PBRF: Phase 1 evaluation - July 2004: Phase 1 evaluation of the implementation of the PBRF and the conduct of the 2003 quality evaluation*. Centre for Research on Work, Education, and Business Ltd. Wellington, New Zealand.

Willmott, H. (2011) Journal list fetishism and the perversion of scholarship: reactivity and the ABS list. *Organisation*, 18: 429-442.

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

VICTORIA UNIVERSITY OF WELLINGTON TE HERENGA WAKA | Chair in Public Finance

Working Papers in Public Finance