



**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 LIMITED**

Wellington Office

Level 13, St John House
114 The Terrace

PO Box 2855, Wellington, New Zealand

Phone: +64-04-499 8140 Fax: +64-04-499 7301

Email: ken.wilson@webresearch.co.nz

Christchurch Office

PO Box 238, Lincoln University

Lincoln, Canterbury, New Zealand

Phone: +64-03-325 3891 Fax: +64-03-325 3829

Email: roberta.hill@webresearch.co.nz



Contents

List of tables.....	4
Preface	5
Executive summary	6
<hr/>	
Introduction	6
Summary of proposals and recommendations	9
Detailed proposals	12
Detailed fundamental recommendations	17
Detailed recommendations for improvements and refinements	21
<hr/>	
Chapter 1: Background to this evaluation	32
<hr/>	
Introduction	32
WEB Research evaluation questions	35
The nature and status of this evaluation	37
Summary	38
<hr/>	
Chapter 2: The individual unit of assessment and the reporting framework.....	42
<hr/>	
Why the design decision was made	42
The arguments for an individual unit of assessment	43
The arguments against the individual unit of assessment	46
Unintended consequences	48
The reporting framework	49
Summary	50
<hr/>	
Chapter 3: Eligibility of individuals and TEOs.....	52
<hr/>	
Staff participation - criteria and definitions	52
Implementation of the Guidelines for determining PBRF staff eligibility	54
The substantiveness test	60
Implementation of the Validation and Verification of Staff Eligibility Audit	64
Summary	66
Eligibility of TEOs	68
<hr/>	
Chapter 4: Direct and indirect costs of the implementation of the PBRF and the conduct of the 2003 Quality Evaluation.....	72
<hr/>	
Introduction	72



Estimating the costs of implementing the PBRF, conducting the 2003 Quality Evaluation and administering the PBRF for the period 2004-2006	75
Estimating the costs of conducting the 2006 Quality Evaluation and administering the PBRF in the period 2007-2012	82
Establishing a possible benchmark for the level of administrative and compliance costs	88
What did TEOs report about their compliance costs in 2003 and their expectations for 2006?	93
Options for savings	106
Chapter 5: Assessing Evidence Portfolios	112
Introduction	112
The status of the findings	115
The structure and composition of the peer review panels	115
The phases of the 2003 Quality Evaluation	117
The preparation, assessment and submission of EPs in TEOs	118
The TEC reception and validation of EPs	124
The assignment of EPs to peer review panels	127
The assessment of EPs by panel members and the analysis by the TEC	131
The assessment of EPs by peer review panels	143
The moderation process	150
The recommendation of Quality Categories to the TEC Board	152
Conclusion	152
References	154
Appendix 1: Phase 1 Methodology	156
Appendix 2: On-line survey of Peer Review Panel Members and Panel Chairs	164
Appendix 3: Abbreviations	216
Appendix 4: Glossary of terms.....	217
Appendix 5: Reports reviewed	220
Appendix 6: Interviews, panel meetings and other events attended	222



List of tables

Table 1:	Estimated Operating Expenditure for the implementation, operation and review of the PBRF for the MOE and the TEC for the period 2002 to 2007.	75
Table 2:	Estimated costs of implementing the PBRF and participating in the 2003 Quality Evaluation to TEOs for the period 2003-2004.....	77
Table 3:	Estimated costs of the implementation of the PBRF and the conduct of the 2003 Quality Evaluation to the MOE, the TEC and TEOs.....	78
Table 4:	Estimated total costs of implementing the PBRF, conducting the 2003 Quality Evaluation and administering and complying with the PBRF in the transition period 2004-2006 to the MOE, the TEC and TEOs.....	81
Table 5:	Estimated costs of administering and complying with the 2006 Quality Evaluation to the MOE, the TEC and TEOs.....	83
Table 6:	Estimated total costs of conducting the 2006 Quality Evaluation and administering and complying with the PBRF in the period 2007-2012 to the MOE, the TEC and TEOs	85
Table 7:	Estimated average dollar return to nine TEOs for each estimated dollar spent on their participation in the PBRF and compliance with the 2003 and 2006 Quality Evaluations.	86
Table 8:	Estimated costs as percentage of total PBRF funding received by nine TEOs for the period 2004-2006 and the period 2007-2012.....	88
Table 9:	Estimated average cost per PBRF-eligible staff member for 2003-2006 for nine TEOs.....	98



Preface

- 1 This report describes and evaluates the impact that the implementation of the Performance-Based Research Fund (PBRF) in 2003 and the conduct of the 2003 Quality Evaluation had upon the Ministry of Education (MOE), Tertiary Education Commission (TEC) and Tertiary Education Organisations (TEOs).
- 2 This report contributes to Phase 1 of *The Performance-Based Research Fund Evaluation Strategy Proposed by the Ministry of Education and the Tertiary Education Commission ('The Evaluation Strategy')* (Tertiary Education Advisory Commission and the Ministry of Education, 2003). Phase 1 focused this evaluation upon early indicators of the impacts of the design and implementation of the PBRF and, in particular, the 2003 Quality Evaluation.
- 3 TEOs that contributed to the study were supportive of the PBRF. Their interest was to improve the PBRF rather than to remove it. They believed that the PBRF needed at least two rounds before judgements could be made about its effects or policy achievements.
- 4 There is willingness in the sector and by the MOE and TEC to review the initial PBRF round in a constructive process. While preserving the integrity and independence of the evaluation, we have sought to contribute to that constructive process.
- 5 Our research design and research processes were developmental. This required us to gather and report on the experiences of the many participants and to reflect those different voices in this report.
- 6 This report is an evaluation report. It is not a detailed analysis of the elements of the PBRF and their rationale, and it is not a report that an expert in performance-based research funds might write as a prelude to the redesign of the PBRF.
- 7 Our evaluation brief and our capacity were limited by the time and budget available to conduct this evaluation. It was also limited by our initial unfamiliarity with the policy, its design elements, their relationships, and their overall strategic significance for TEOs and individuals.
- 8 We could not have received better co-operation from any organisation or individual participant. The WEB Research evaluation team of Phillip Capper, Ken Wilson and Dr. Roberta Hill, and our Associates, Tony Bullard and Suzanne Vallance, thank the very large number of people who contributed to this report. We also thank members of the Evaluation Advisory Group (EAG) for assisting us with the development of the argument in this report.

Executive summary

Introduction

The Evaluation Strategy

- 9 The MOE and TEC proposed an Evaluation Strategy for the PBRF that was issued for sector consultation in August 2003. The Evaluation Strategy has three phases. Phase 1 focuses upon the design and implementation of the 2003 Quality Evaluation, in particular:
 - an evaluation of the implementation process (especially in relation to the 2003 Quality Evaluation);
 - the short-term impacts of the PBRF on the tertiary education sector, including modelling the likely financial implications of the PBRF for TEOs during 2004-2007; and
 - the results of the Quality Evaluation and what these reveal about the overall quality of research being conducted in the tertiary education sector, the main areas of research strength and weakness, and the relative research performance of the TEOs that have participated in the PBRF (Tertiary Education Advisory Commission and the Ministry of Education, 2003, p5).
- 10 Phase 2, the medium-term phase, will focus on a more detailed review and evaluation of the wider impacts of the PBRF on the tertiary education sector.
- 11 Phase 3, the longer-term phase, will focus on whether the PBRF has fulfilled its stated objectives and whether the overall benefits have exceeded the costs. (Phase 3 will be undertaken after the second Quality Evaluation but prior to the third Quality Evaluation due in 2012).
- 12 Those parts of the Phase 1 Evaluation Strategy, not covered in this evaluation as described in the Preface, are being carried out by the MOE and TEC. These include:
 - changes to transition requirements;
 - the need for buffering of funding to providers;
 - the impacts of the research degree completions (RDC) measure on taught degree programmes and research quality;
 - the financial viability of particular TEOs; and



- the results of the 2003 Quality Evaluation. These were reported in *Performance-Based Research Fund: Evaluating Research Excellence: the 2003 assessment ('The 2003 assessment')* (Tertiary Education Commission, 2004a).

Evidence for this evaluation

- 13 This evaluation has been carried out by the Centre for Research on Work, Education and Business Ltd (WEB Research). The evaluation design for this report is set out in Appendix 1. This evaluation report forms part of the MOE and the TEC Phase 1 Evaluation Strategy. To compile this report, we have drawn upon the following kinds of evidence:
- in-depth interviews with staff of the TEC (especially the PBRF Project Team) and the MOE;
 - interviews with individuals or groups representing key stakeholder organisations;
 - observation of the Moderation Panel and a sample of peer review panels;
 - interviews with panel chairs and members;
 - a survey of peer review panel chairs and members;
 - scoping interviews in two TEOs;
 - case studies in three TEOs;
 - scrutiny of TEC documentation relating to the design and administration of the PBRF;
 - iterative consideration of emergent themes with members of an Evaluation Advisory Group (EAG);
 - use of incidentally occurring data such as public statements to the media, academic papers, and meetings or workshops; and
 - publications and documents relating to jurisdictions referred to in designing the PBRF, especially the experiences of similar performance-based funds in the research assessment exercises (RAEs) conducted in the United Kingdom (UK RAE) and Hong Kong.

Outline of this evaluation

- 14 The implementation of the PBRF, including the conduct of the 2003 Quality Evaluation, was carried out successfully despite a very tight

timeline. The rapid implementation revealed problems, but in our judgement none of these is an immediate and fatal threat to the policy. Some problems are more serious and represent a risk to the longer-term policy goals of the PBRF. Some of these matters require attention before the next Quality Evaluation (scheduled for 2006) takes place.

- 15 No TEO or other stakeholders consulted in this evaluation advocated any fundamental re-design of the PBRF or the Quality Evaluation process. On the contrary, there was a broad expectation that the 2006 Quality Evaluation would remain more or less the same and that the on-going operation of the PBRF would proceed with continuous improvements. This evaluation points particularly to design and process improvements to the Quality Evaluation process that could be made without imposing additional costs on TEOs.
- 16 The implementation of the PBRF and the conduct of the 2003 Quality Evaluation were possible because of a high level of trust and co-operation between TEOs, their staff and the TEC. There is a very strong expectation in TEOs that the TEC will continue to develop that trust and will maintain good relationships with the sector. There is an equally strong expectation that the guidelines for 2006, incorporating all changes to rules, processes, scoring, data returns and audits, will be with TEOs in time for the beginning of the 2005 academic year.
- 17 Overall, the evaluation suggests that TEC attention should focus on modifying the existing processes in time for the less hurried conduct of the 2006 Quality Evaluation. This evaluation does not suggest that attention be given to a fundamental redesign of the PBRF or the Quality Evaluation for three reasons:
 - there was no evidence of major design failure reported or observed in the evaluation;
 - while participants identified problems with design elements and aspects of the implementation, these were not seen as immediate and fatal threats but rather as remedial; and
 - it is too soon to observe or evaluate the impact of the PBRF upon TEOs and individuals.
- 18 Participants in this evaluation of the 2003 Quality Evaluation understood that it is not possible to devise a completely objective, or error-free, qualitative evaluation of research. They also understood the considerable difficulty in putting such a scheme into operation.
- 19 Their expectation is that the broadest consensus will be developed around: what the standards are to be; how they are understood within the TEC and TEOs: and how they are to be interpreted by the peer review panels. The focus, between Quality Evaluations, should be on

continuous improvement to the fairness and consistency of the practice of the Quality Evaluations.

- 20 A large number of issues emerged in the course of the evaluation. Many are of a minor nature and are likely to be corrected by small adjustments and modifications. Others present greater problems. Those are issues near the heart of the design of the PBRF and they have typically resisted correction in other jurisdictions. There are opportunities to mitigate their effects and minimise their negative impacts. For the MOE, TEC and TEOs, improving and sustaining the PBRF is a question of managing these issues and making trade-offs and compromises where necessary.

Structure of the Executive Summary

- 21 First we present a summary of our proposals and recommendations. In the sections that follow we provide a detailed examination of those proposals and recommendations. We distinguish these in the following way:
- our **proposals** concern what should be retained from the basic design and from existing procedures;
 - our **fundamental recommendations** relate to significant modifications to address significant risks to the policy; and
 - our **recommendations for improvements and refinements** are modifications required if the main design elements of the PBRF are retained.

Summary of proposals and recommendations

- 22 In the UK, the *Review of Research Assessment* ('*Review*') noted that there is a dual purpose in assessing research – the allocation of funding and the provision of high-quality information (Roberts, 2003). The *Review* explains that:
- these dual purposes impose intense stresses and costs on the quality-assessment process because they demand reporting information in two different ways; and
 - information for the public and policy makers needs to be traded off against the higher need to serve the fair and transparent allocation of funding.
- 23 The PBRF has the same dual purpose. Where there is an apparent tension between the two, the proposals and recommendations in this report have given precedence to the need to provide a fair and

transparent funding mechanism. This is because the evaluation showed that sector confidence in the PBRF needs to be fostered. Giving priority to demonstrating fairness and transparency, rather than to the information needs of the public or policy makers, is more likely to sustain the PBRF during its transition years.

- 24 We **propose** that the following features of the 2003 Quality Evaluation should be retained in the next Quality Evaluation (be that 2006, or later):
- a mixed model of peer review and quantitative performance indicators, using the three components of the 2003 Quality Evaluation – that is, the Quality Evaluation, the research degree completions (RDC) measure, and the external research income (ERI) measure;
 - the individual staff member as the unit of assessment in the Quality Evaluation;
 - quality assessment carried out by multidisciplinary peer review panels whose membership is drawn from both New Zealand and overseas;
 - the principle of a standards-based holistic assessment of research quality that includes research output (RO), peer assessment (PE), and contribution to research environment (CRE) components;
 - audit and verification processes for both staff eligibility and the evidence included in Evidence Portfolios (EPs);
 - the reporting framework for the Quality Evaluation (with the proviso that it be subject to a consultative review to consider the privacy and confidentiality of individual Quality Categories); and
 - the framework used for reporting the results of the PBRF to TEOs and the public (that is, only by aggregation of results and not by individual Quality Categories).
- 25 These features form the foundation of a robust, transparent and defensible funding-allocation process.
- 26 If the proposals to retain the central design elements of the 2003 Quality Evaluation are accepted, there are a number of adjustments that need to be made and conditions that need to be met in order to strengthen them.
- 27 The following set of recommendations address those adjustments. We **recommend** that:



- The criteria for determining staff participation in the PBRF be reviewed and revised. In particular:
 - i. the intended coverage of the Staff Participation Criteria (employment criteria) be clarified;
 - ii. the substantiveness test be reviewed for its scope and application; and
 - iii. the eligibility audit be a more systematic and robust examination of a TEO's application of the eligibility criteria.
- The timetable for the 2006 Quality Evaluation provide for:
 - i. a consultative review process between the TEC and the sector; and
 - ii. a commitment by the TEC to develop an improved IT platform that will facilitate the creation, collection and processing of EPs.

If these matters cannot be implemented in time to enable the 2006 Quality Evaluation to be conducted in a timely manner, we suggest that the TEC consider approaching Cabinet and seeking authority to delay the 2006 Quality Evaluation.

- The administrative and compliance costs of the PBRF and 2006 Quality Evaluation for the MOE, TEC, and TEOs be no more than 2% of the total PBRF funding allocated for the period 2007 to 2012.

28 If the proposals to retain the main design elements of the 2003 Quality Evaluation are accepted, there are a number of process improvements that should be instituted to improve the fairness, consistency and robustness of the Quality Evaluation. We **recommend** that:

- TEOs be required to conduct an assessment of all PBRF-eligible staff and provide Quality Category information to the TEC in 2006;
- the current number of peer review panels be retained, but their structure, subject composition and selection processes be reviewed;
- the peer review panels be enjoined to make greater use of specialist advice and cross-referral where an EP or a specific NRO relates to a discipline or sub-discipline not represented on a particular panel;
- the interpretation and application of the PBRF definition of research (as amended) be re-examined in Phase 2 of the Evaluation Strategy;

- the guidelines for assessing EPs be modified by the application of the experiences of the 2003 peer review panels, as recorded in their reports and in this evaluation;
- the tie-point descriptors of quality be revised to more effectively identify the quality of applied research, New Zealand research, and practice-based research;
- the scoring system be amended to eliminate or mitigate the disadvantages and distortions affecting certain groups as described in this evaluation; and
- the TEC:
 - i. develop an improved IT platform that will facilitate the creation, collection and processing of EPs, that will be effective across the range of platforms used by TEOs, that will minimise errors in data presentation, and that will minimise the costs of providing and maintaining PBRF data for TEOs; and
 - ii. advise the sector of its intention to develop such a platform as soon as practical, so that TEOs can plan to integrate this platform into their internal systems; and that
- the suggestions of the National Library of New Zealand on the auditing of NROs be used as a basis for refining the EP compilation process.

Detailed proposals

The funding model

Proposal 1

That a mixed model of peer review and quantitative performance indicators using the three components of the 2003 Quality Evaluation (that is, the Quality Evaluation, an RDC measure, and an ERI measure) be retained.

- 29 The PBRF Working Group recommended a funding model with three components: a periodic 'quality evaluation' of research excellence, achieved by peer review; an RDC measure; and a measure for ERI. These three components produce a funding formula for TEOs in the proportion 60:25:15. This is known as the 'mixed' model (Ministry of Education and Transition Tertiary Education Commission, 2002, p10).
- 30 The use of a peer review component (by contrast with Australia, for example, where a quantitative measure of research output is used) avoids many of the perverse effects associated with equating quantity of

outputs with quality, and was affirmed by those academics who had experience of both approaches.

- 31 The primary arguments against peer review are its high cost and complexity. We have found that the complexity of the peer review process has been well managed, and that the administrative and compliance costs of the initial Quality Evaluation were high. There are also costs involved in altering the model in significant ways after only one Quality Evaluation, without there being any certainty that the change would be an improvement. We therefore propose that the peer review component be retained.

The individual as the unit of assessment

Proposal 2

That the individual staff member be retained as the unit of assessment in the Quality Evaluation.

- 32 The PBRF Working Group considered a number of arguments for selecting the individual staff member as the unit of assessment for the Quality Evaluation. The PBRF Working Group saw the small size of the New Zealand academic community as meaning that having discipline-based panels would be difficult and expensive so it had to accept multidisciplinary peer review panels. As multidisciplinary peer review panels work more effectively if the unit of assessment is the individual, the PBRF Working Group accepted that the unit of analysis would be the individual. The rationale for choosing the individual as the unit of assessment is being reviewed on a number of grounds, with confidentiality and cost being the most significant.
- 33 This evaluation suggests that confidentiality issues are unlikely ever to be eliminated, and that moving to a group-based unit of assessment may not reduce them either. While the costs of using an individual unit of assessment were expected to be greater than in other performance-based assessment methods, a reliable estimate of the level of those costs will only be possible when the PBRF is fully implemented. An estimate of the possible levels of costs for the 2006 Quality Evaluation and the PBRF in the period 2007-2012 suggests that the costs of using the individual unit of assessment, in the New Zealand context, may approach the costs of the assessments used in other comparable performance-based funding-allocation systems.
- 34 We have earlier expressed the view that it is undesirable to change the assessment framework fundamentally after only one Quality Evaluation. There would need to be very compelling reasons to do so and there would need to be a compelling case that the change would represent an improvement. We believe that both the evidence so far and the

alternatives available, do not make a compelling case for a fundamental design change relating to the unit of assessment.

- 35 This is not to say that, in the medium to long term, issues of confidentiality do not represent a significant threat that might compel more radical modification of the Quality Evaluation.

Multidisciplinary peer review panels

Proposal 3

That the quality assessment be carried out by multidisciplinary peer review panels whose membership is drawn from both New Zealand and overseas.

- 36 The decision to use multidisciplinary panels rather than a larger number of discipline-based panels was originally seen as a forced choice primarily associated with the small size of the New Zealand tertiary education sector. Our reasons for proposing that peer review panels continue to be multidisciplinary are broadly the same as for the retention of the individual as the unit of assessment.

The Quality Evaluation assessment framework

Proposal 4

That the principle of a standards-based holistic assessment of research quality that includes RO, PE and CRE components be retained.

- 37 The PBRF Working Group argued that being an excellent researcher included a range of activities other than 'just... the production of well-respected articles, books and other forms of research output', and therefore recommended that the Quality Evaluation should 'assess the overall performance of researchers', using RO, PE and CRE (Ministry of Education and Transition Tertiary Education Commission, 2002, p7 & p13).
- 38 Although few Quality Categories were directly changed as a result of the overarching holistic assessment, the provision for an holistic assessment underpinned and informed the scoring process in many panels. The holistic assessment is a significant signal that the Quality Evaluation is based upon explicit quality standards and criteria, and that it is concerned with peer review of research quality not quantity. Although the holistic assessment is not an essential component of a standards-based assessment, we propose that it be retained.



Audit and verification

Proposal 5

That the audit and verification processes for PBRF staff eligibility and EPs continue.

- 39 Rigorous audit and verification processes are essential for ensuring fairness and robustness especially while confidence in the system develops. TEOs identified issues associated with the various audit processes that will inform any review of the future scope and nature of those audits. Improving the audit of staff eligibility and the application of the substantiveness test is particularly important.

The reporting framework

Privacy and confidentiality

Proposal 6

That the reporting framework for the Quality Evaluation be retained but be the subject of a consultative review that considers the privacy and confidentiality of individual Quality Categories.

- 40 There are significant privacy and confidentiality issues involved in using the individual as the unit of assessment, and this matter was considered in *Investing in Excellence: The Report of the Performance-Based Research Fund Working Group* (*Investing in Excellence*) (Ministry of Education and Transition Tertiary Education Commission, 2002, p11). The TEC remained sensitive to this issue and in *The 2003 assessment* describes the many attempts it took to achieve a reporting framework that would maintain individual confidentiality (Tertiary Education Commission, 2004a, p36). The TEC resolved to hold to the original reporting framework even though this made it possible for individual Quality Categories to be inferred in many cases.
- 41 The smallness of New Zealand's academic community, and especially the small numbers of eligible staff in some departments and subject areas, has made it easy in many cases to infer the Quality Categories assigned to individual staff. This has produced some distress and disquiet; it may have a negative effect on sector attitudes towards the PBRF, and on decisions made by individual staff concerning their own careers.
- 42 Privacy and confidentiality problems appear to be serious and not readily solvable. This has led some members of the PBRF Project Team, who have been closely involved in its implementation, to reconsider the decision to use an individual, rather than a group, unit of assessment for the Quality Evaluation. Yet even such a change seems unlikely to solve the problem given:

- the smallness of the New Zealand academic community overall and the smallness of many academic units within it;
 - the framework of administrative law in New Zealand (especially the provisions of the Privacy Act and the Official Information Act); and
 - the framework of employment law in New Zealand (including collective bargaining and the existence of experienced, organised labour).
- 43 In such circumstances we propose that it is better to maintain the status quo in terms of assessment units, and to engage all involved in open dialogue about how improvements might be made. The employment law framework and the existence of experienced and organised labour provide opponents of the PBRF with a legal mechanism for undermining the PBRF. This is not the case to the same extent legally or culturally in the UK or Hong Kong. An open dialogue with stakeholders would allow long-term solutions to be considered from every viewpoint and for all stakeholders to develop an understanding of the behaviours that need to be adopted for the PBRF to gain the continuing support of all parties. This matter needs to be revisited and reconsidered after future Quality Evaluations.
- 44 While the privacy and confidentiality of individual Quality Categories can never be guaranteed, breaches of confidentiality and abuses that might occur when confidentiality is breached can be mitigated by improved procedures and by the adoption of codes of practice and acceptable-use codes. Such codes cannot mandate individual or institutional behaviour, but they do have moral force if collaboratively designed.
- 45 TEOs were allowed to nominate the academic units into which Quality Categories were to be grouped for reporting purposes (Tertiary Education Commission, 2004a, p37). This had two effects:
- some TEOs chose to define relatively small academic units despite a prior warning from the TEC that this would make it easier to infer individual Quality Categories; and
 - the range of decisions made by TEOs made it impossible to reliably compare the relative performance of nominated academic units within similar disciplinary areas. The use of subject-area quality scores in the reporting the results helped to overcome this problem; but this caused other problems such as apparent distortions arising from the groupings chosen.

Classification of subjects

Proposal 6

That the framework for reporting PBRF results to TEOs and the public continue to be by the aggregation of Quality Categories.

- 46 Prior to the 2003 Quality Evaluation, the TEC developed 41 groupings of subjects for reporting purposes. This was done in consultation with the tertiary sector. In doing this, the TEC was conscious that decisions on this matter involved a trade-off:
- the more that subject results could be disaggregated and results reported discretely, the more valuable would be the information and the more valid would be comparisons; but
 - the greater the degree of disaggregation, the greater the capacity for an interested outsider to infer the Quality Categories assigned to individual staff members.
- 47 The implementation of a reporting framework based upon the 41 subject areas chosen has created disquiet in some academic units in TEOs. An example of this is where a strong research subject has been aggregated with one or more weaker subjects, with the result that the apparent performance of the strong subject has been concealed or diluted. Decisions made in different TEOs with respect to the structure of their nominated academic units has exacerbated or mitigated this problem locally.
- 48 Addressing this effect by significantly disaggregating the current subject areas would greatly increase the risk of loss of confidentiality. Major changes would also reduce the capacity to make valid comparisons over time. Nevertheless there may be some opportunities for improvement in the groupings that would minimise the possibility of such negative consequences.
- 49 In the body of the report, we make suggestions that are intended to assist in any review of the subject groupings.

Detailed fundamental recommendations

- 50 These recommendations relate to fundamental design issues that, unless they are resolved or significant progress is made towards resolving them, represent a significant threat to the success of the PBRF as a funding mechanism.

Eligibility criteria (staff and TEOs)

Recommendation 1

The criteria for determining staff participation in the PBRF be reviewed and revised. In particular:

- *the intended coverage of the Staff Participation Criteria (employment criteria) be clarified;*

- *the substantiveness test be reviewed for its scope and application; and*
 - *the eligibility audit be a more systematic and robust examination of a TEO's application of the eligibility criteria.*
- 51 For TEOs in the 2003 Quality Evaluation, identifying PBRF-eligible staff was both a compliance task and a strategic task:
- TEOs were required to identify PBRF-eligible staff in accordance with the 'Performance-Based Research Fund: A Guide for 2003' (the *Guidelines*) (Tertiary Education Commission, 2003) and obtain EPs from those staff; and
 - TEOs were in a position to influence their overall reported quality score by the way they chose to interpret and apply the criteria in the *Guidelines* and particularly how they applied the substantiveness test.
- 52 Determining the total number of PBRF-eligible staff is critical for a TEO because all PBRF-eligible staff are included in the denominator for calculating the quality score. However, under the current policy framework, a TEO faces conflicting pressures in applying the eligibility guidelines. On the one hand, it has an incentive to minimise the number of eligible staff (especially staff likely to secure an 'R') in the interests of reducing the size of its denominator and thus increasing its likely quality score. On the other hand, if it excludes staff members who have a chance of securing at least a 'C', it runs the risk of reducing the funding to which it is entitled.
- 53 There have been major changes to the employment framework in New Zealand following the Employment Contracts Act (1991) and the Employment Relations Act (2000). These Acts permit and enable a variety of forms of employment relationships that, as TEOs reported, were not captured by the definition of employment in the eligibility criteria. Staff who are not employed on a permanent full-time basis may be employed as casual, casual part-time, permanent part-time, and even permanent casual part-time.
- 54 The speed of the implementation process and the unfamiliarity of TEOs with the Staff Participation Criteria did not allow for informed fine-tuning of the eligibility criteria during the 2003 Quality Evaluation. There was no time to ensure that the criteria were workable at the level of a TEO or that they were properly and consistently applied.
- 55 The eligibility audit was a mechanism to provide:
- assurance to the TEC that the appropriate staff had been included in the 2003 Quality Evaluation; and

- assurance to TEOs that all TEOs were entering the 2003 Quality Evaluation on an equal basis – especially as the eligibility criteria were new, open to interpretation, and perceived to be capable of being manipulated.
- 56 TEOs and stakeholders in this evaluation were strongly of the view that the eligibility rules must be seen to be applied consistently by all participating TEOs. The PBRF Staff Eligibility Audit is both an assurance that the rules are being applied consistently and a transparent, public base for creating and maintaining confidence in the whole system.
- 57 The perception of TEOs in this evaluation was that the eligibility audit was not an adequate audit of the eligibility criteria nor an adequate audit of the TEOs' application of the substantiveness test.
- 58 Within the limits of this evaluation, we observed no examples of creative or mischievous interpretations of the eligibility criteria.
- 59 The PBRF was intended to integrate a funding allocation mechanism with a measure of research quality so as to stimulate excellence in research and lively research environments across the tertiary sector. The PBRF provided an opportunity for all TEOs to have the excellence of their research assessed, but not all TEOs participated in the 2003 Quality Evaluation.
- 60 If the overall intent of the PBRF in its 2003 Quality Evaluation was 'to reward and encourage excellence [in research]' in all TEOs, then the responses of most polytechnics, some wānanga, and some private training enterprises were not congruent with that objective as they elected not to participate. This does not mean that the policy will not succeed over time. It may be that TEOs which wish to secure funding from the PBRF in the future will elect to participate at that time. And it may be that TEOs which did not participate in the 2003 Quality Evaluation nevertheless intend to prepare their staff to participate in a future Quality Evaluation.
- 61 As was the case in 2003, the main participants in the 2006 Quality Evaluation appear likely to be the universities. As universities currently contain most of the PBRF-eligible staff within the tertiary education sector, this will mean that the policy intent will largely be met.

Administrative infrastructure

Recommendation 2

That the timetable for the 2006 Quality Evaluation provide for:

- *a consultative review process between the TEC and the sector; and*



- *a commitment by the TEC to develop an improved IT platform that will facilitate the creation, collection and processing of Evidence Portfolios.*
- 62 Many stakeholders have either told us, announced publicly, or included in their reports on their involvement with the process that it is desirable to proceed with the 2006 Quality Evaluation as is currently intended. We believe that it is more important that the infrastructure requirements and implementation of modifications should be completed before the next Quality Evaluation. If this means that the 2006 Quality Evaluation is put back, then the TEC should make a decision to do so as soon as possible.
- 63 The implementation of a full cycle of the PBRF, from policy decisions to funding allocations in the period concluding in April 2004, was a remarkable accomplishment by the TEC and the tertiary education sector. The peer review panels and their secretariats worked extremely hard with very high levels of commitment to the process.
- 64 A significant component of the administrative success of the 2003 Quality Evaluation was the excellent relationships that were established between members of the PBRF Project Team and administrative staff in TEOs. Most of the Project Team and some TEO staff were working under fixed-term contracts and few of them have been retained, with a consequent loss of networked relationships and some tacit institutional knowledge.

Direct and indirect costs

Recommendation 3

That the administrative and compliance costs of the PBRF and 2006 Quality Evaluation for the MOE, TEC, and TEOs be no more than 2% of the total PBRF funding allocated for the period 2007 to 2012.

- 65 The estimate of costs to the MOE, TEC, TEOs, and staff indicate that the costs of the implementation of the PBRF and especially the 2003 Quality Evaluation were high as a proportion of:
- the funding allocated on the basis of the 2003 Quality Evaluation; and
 - the total funding that will be allocated by the PBRF in the transition period 2004-2006 (approximately \$170 million).
- 66 The PBRF replaces an existing mechanism for delivering funding to TEOs. The PBRF Working Group recognised that a performance-based funding delivery mechanism would be more expensive (less efficient) than the EFTS-based funding delivery mechanism. But what might be an



appropriate level of administrative and compliance costs for the PBRF could not be known until the PBRF was fully implemented.

- 67 Limited comparisons can be made with the level of administrative and compliance costs reported by other performance-based funds such as the UK and Hong Kong RAEs. Those comparisons help to identify a possible benchmark for administrative and compliance costs of the PBRF in 2007.
- 68 From 2007 onwards the PBRF will be fully implemented, with a total annual allocation of approximately \$185 million (on current estimates). Assuming that the government provides some new funding, the PBRF is likely to allocate around \$1.2 billion in the six years following the 2006 Quality Evaluation. On this basis, the administrative and compliance costs of the PBRF will be much lower as a proportion of the total funds allocated than will be the case during the three-year transition period 2004-2006.
- 69 Estimates of costs supplied by the MOE and some TEOs indicate that the percentage of administrative and compliance costs as a proportion of the total funds allocated is likely to fall from a range between 12% and 17% for the period 2004-2006 to a range between 1.45% and 1.93% for the period 2007-2012.
- 70 This would be a considerable accomplishment and would place the proportion of administrative and compliance costs of the PBRF close to the proportions reported for the UK and Hong Kong RAEs. However, given the many imponderables involved in forecasting costs for 2007 to 2012, a realistic goal may be to ensure that the administrative and compliance costs of the PBRF for the period 2007-2012 (including the costs of the 2006 Quality Evaluation) do not exceed 2% of the total funds allocated for that period.

Detailed recommendations for improvements and refinements

- 71 If the proposals to retain the main design elements of the 2003 Quality Evaluation are accepted, there are a number of process improvements that should be instituted to improve the fairness, consistency and robustness of the Quality Evaluation. The following section discusses those recommendations.

Internal assessment by TEO panels

Recommendation 4

That TEOs be required to conduct an assessment of all PBRF-eligible staff and provide Quality Category information to the TEC in 2006.

- 72 The original design of the Quality Evaluation envisaged the possibility of an increasing degree of self assessment within TEOs, with the external peer review panels being relegated progressively to an auditing role (Ministry of Education and Transition Tertiary Education Commission, 2002, pp14-15). More recently Boston has stated that the 'whole idea of self-assessment was rather ambitious, if not misguided' (Boston, 2004, p7).
- 73 We agree that self-assessment as envisaged is no longer possible. We believe that it is desirable for the external peer review panels to remain as the assessing body. But this evaluation indicates that for TEOs to carry out an internal assessment has proved important in building understanding and capability within TEOs about the quality assessment processes and the standards applied. We consider it desirable that this step be retained for at least one more Quality Evaluation.

Number and composition of peer review panels

Recommendation 5

That the current number of peer review panels be retained, but that their structure, subject composition and selection processes be reviewed.

- 74 The peer review panel process was well-designed and subject to rigorous moderation procedures. Members of panels did more than had been expected of them in pursuing the interests of fairness and robustness. Their work was hampered both by the variability in the quality of the evidence contained in the EPs they were given to work with and by exacting time constraints.
- 75 However, the structure and composition of the panels was problematic. The fundamental decision to adopt a small number of multidisciplinary panels (like Hong Kong) rather than a larger number of disciplinary panels (like the UK) was considered to be inevitable for a small country. This has produced some apparent inconsistencies in assessments of EPs in subject areas where there was no person from that subject area on the panel, or where one disciplinary area was assessed by more than one panel.
- 76 There are two factors to consider with respect to the peer review panels: the number of panels and the subject areas they were assigned; and the representativeness of the panels. The latter has two aspects - the differing sizes of panels; and the proportion of members of a panel who were specialists in each of the disciplines assigned to it.
- 77 The number of panels was broadly a function of the decision to use multidisciplinary panels. We see no imperative to change that decision and there are considerable potential problems in doing so, particularly those of logistics and costs. The exact number of panels is not in itself

significant and could be changed. However, we believe that the best approach after one round is to keep the number of panels stable, but to adjust the assignment of subjects to each one to achieve a better workload balance if that is possible. Suggestions to adjust the assignment of subject areas to panels are discussed in Chapter 5.

- 78 There were workload imbalances between panel members. This was because the subject-area expertise in panels did not match the subject areas of the EPs allocated to panels evenly. A close analysis of the 2003 panel expertise and EP subject areas would allow better judgements to be made about future panel membership.
- 79 There are opportunities to improve the peer review process by modifying the composition of panels. In the 2003 Quality Evaluation, some subjects or sub-disciplines were over-represented on panels while others were under-represented. For example, four law panel members who experts in public law, while no panel member had expertise in commercial law.
- 80 Although the TEC produced an internal report on the panel appointments process, there was no external audit of panel representativeness. An external audit could have provided assurance of the fairness of the selection process and we suggest this possibility be considered.
- 81 The frequency of declared conflicts of interest, inevitable because of the small size and close relationships within the New Zealand academic community, added complexity to the panel processes. This is an inherent characteristic for which there is no permanent solution though greater use of specialist advice may mitigate its impact.

Use of specialist advice and cross-referral by peer review panels

Recommendation 6

That peer review panels be enjoined to make greater use of specialist advice and cross-referral where an EP or a specific NRO contains elements relating to a discipline or sub-discipline not represented on a particular panel.

- 82 This is a corollary of recommendation 5.
- 83 Under the *Guidelines*, a panel could obtain additional input into the assessment of an EP when the members of the panel could not provide all the expertise necessary to fully review an EP. There were two sources of additional input:
- another panel (cross-referral); and
 - a specialist adviser.

- 84 The *Guidelines* provided that peer review panels could use specialist advisers if:
- relevant subject-area expertise was not represented on the panel;
 - conflict of interest ruled out the use of a panel member's subject-area expertise; and
 - members of a panel with the relevant subject-area expertise could not reach a consensus on the scoring of components of an EP.
- 85 Panels referred 87 EPs to specialist advisers.
- 86 The *Guidelines* provided for input from another peer review panel 'typically' when a significant proportion (but not a majority) of a cross-disciplinary RO fell within the subject areas covered by another panel. Panels cross-referred 485 EPs for this reason.
- 87 This evaluation indicates that not all EPs or nominated research outputs (NROs) that might have justified obtaining additional input under the *Guidelines* were sent on for such input. We are not able to quantify the extent to which this was so, but draw on evidence from:
- observations recorded in some panel reports;
 - our own observations of panels at work;
 - comments made to us by some panel members in interview;
 - comments from panel members in the survey conducted as part of this evaluation; and
 - interview data from the TEC and panel secretariats.
- 88 Our observations of instances where specialist advice was sought and cross-referrals were made by panels, suggested that the facility to obtain additional input was used carefully by the panels – although there was some inconsistency between panels. Further and more detailed evaluation of the conduct of panel assessments might lead to guidelines to help achieve greater consistency both in panels' decisions to use additional input and in the use made of such input.

Definition of research

Recommendation 7

That the interpretation and application of the PBRF definition of research (as amended) be re-examined in Phase 2 of the Evaluation Strategy.

- 89 The PBRF Working Group adopted a 'new definition of research, which is specific, yet wide ranging and enables excellence to be recognised

wherever it occurs' (Ministry of Education and Transition Tertiary Education Commission, 2002, p10).

- 90 This definition arguably formalises what exists as tacit practice in universities. It is largely based on the definitions used in the UK RAE, the New Zealand Qualifications Authority (NZQA), and the OECD. However, it is not identical to the NZQA definition; and for non-university TEOs this was a critical issue. In addition, for some subjects there remains a question as to whether the definition itself, or its qualitative interpretation by peer review panels, did in fact adequately recognise excellence wherever it occurred.
- 91 The issues relating to the interpretation of the definition of research are complex, and not susceptible to simple solutions. They include:
- the interpretation of its meaning by the peer review panels, and especially its application to a continuum of activities where many fall ambiguously on the boundary of what is included and what is not;
 - the problem of applying the definition in 'non-traditional' research areas, such as the creative and performing arts; and
 - the question of whether modifiers (such as the following) might be better included in the definition itself:

The quality evaluation process will give full recognition to work of direct relevance to the needs of industry and commerce, and all research, whether applied or basic/strategic, will be given equal weight (Ministry of Education and Transition Tertiary Education Commission, 2002, p18).

Assessing the Evidence

Recommendation 8

That the guidelines for assessing EPs be modified by the application of the experiences of the 2003 peer review panels, as recorded in their reports and in this evaluation.

- 92 Eligible staff participating in the 2003 Quality Evaluation were required to submit EPs that included RO, PE and CRE. Eligible staff could submit up to 54 ROs; and up to four of these could be NROs (that is, 'nominated' as representing the staff member's best work). NROs had to be available for examination by panels on request.
- 93 It was quite impractical for members of panels to sight all NROs within the available timeframe for the assessment process. The *Guidelines* required them to sight at least 10% of NROs (Tertiary Education Commission, 2003, p199), but in practice most panel members sighted

at least 20% (Tertiary Education Commission, 2004a, p27) and felt obligated to do so in the interests of fairness. For the remainder of the NROs, panel members necessarily used proxy indicators contained within the EP.

- 94 The *Guidelines* emphasised that panel members were not expected to undertake as full an examination of an NRO as they would if they were conducting a formal peer review, but rather that the purpose was to assist in forming a judgement about the overall quality of the RO component of an EP.
- 95 Assessing the CRE and PE components of EPs was largely dependent on the clarity and content of the evidence included by the staff member preparing the EP. This evidence was judged against tie-point descriptors provided in the *Guidelines*.
- 96 In practice, the examination of NROs often became a proxy indicator for the whole of the RO component of EPs. The PBRF assessment process has no choice other than to use proxy indicators of quality. It is impractical to conduct a full and formal peer review of all the evidence provided. The validity of the judgements made were therefore dependent on:
 - how well the evidence in an EP was presented and documented; and
 - the consistent and appropriate use of the proxy indicators by panel members.
- 97 The most commonly used proxy indicator of quality by peer review panels was the international standing of journals. In our observations of the panel processes, the standing of a journal was judged either by impact factors, or by the 'rule-of-thumb' opinion of subject specialists on the panel. These judgements about journals were a central component of the assessment process, although mitigated and set alongside judgements made about the NROs that were examined.
- 98 A major difficulty was that, while publication in a leading journal could be taken as evidence of world-class research, panel members were often confronted with the problem that publication in a lesser journal, or in a New Zealand or regional journal, could not necessarily be taken to indicate lower quality.
- 99 In the absence of clear positive evidence of quality, we observed panel members often scored an NRO as though the journal in which it was published represented its quality.
- 100 While using the standing of journals as measured by impact factors is itself a difficulty, panels confronted greater difficulties in establishing consistent and appropriate proxy measures for making judgements about other types of RO such as presentations, performances and

artistic works. In practice we noted, or were told of, instances where panel members found it difficult to judge whether or not ROs were, in fact, quality assured, or what degree of quality could be inferred from refereeing or review processes.

- 101 Significant issues such as those above, need to be addressed with respect to the use of proxy measures. In any process that involves making judgements about quality there will always be a large element of subjectivity.
- 102 We consider it desirable that the Quality Evaluation continues to concentrate on measures of quality, and not compromise doing so because the process is difficult. To compromise would involve using more quantitative measures, such as journal impact factors and quantity of ROs.
- 103 There is a difficult trade-off to be made. Retaining an holistic assessment with three components results in greater subjectivity in the process and therefore increases the risk of perceived unfairness. Greater use of quantitative measures would result in more consistent decisions; but it would also result in an impoverished understanding of what constitutes quality in research.
- 104 In our view, it is in the best interests of the sector, the TEC and the intent of the overall PBRF policy to accept the uncertainties associated with the approach to quality assessment used in 2003, and to develop and refine – progressively - the understanding of how to improve the use of proxy evidence. If all parties fully engage in such a developmental process, it is more likely that the process will be accepted despite its inevitable imperfections.
- 105 In Chapter 5 we make observations designed to assist such a developmental process.

Tie-point descriptors

Recommendation 9

That the tie-point descriptors be revised to more effectively identify the quality of applied research, New Zealand research, and practice-based research.

- 106 Of all the issues they faced, peer review panel members were least confident that they established robust inter-subject comparability, and they were quite specific in their comments about which types of subject they felt might have been most disadvantaged. Our observational data supports this view. The RAE experience in the UK suggests that this is not an easy problem to deal with.

- 107 The tie-point descriptors provided in the *Guidelines* and used by panel members to assess EPs were comprehensive. However, they contained a number of phrases that proved problematic in their interpretation by members of the panels. The most important of these was the phrase 'world class research' used in the tie-point descriptor applied to RO at level 6 in the 0-7 point scale.
- 108 Although the *Guidelines* for panel members specifically stated that 'research outputs that deal with topics or themes of primarily local, regional or national focus can be of world standard', in practice peer review panel members often found it difficult to judge whether this was so for any particular RO unless it was one that they personally sighted and had personal subject expertise about. Proxy evidence used by panels did not always allow the panel to make an accurate judgement.
- 109 As we note above, the most commonly used proxy indicator of quality was the international standing of journals, although publication in a lesser journal, or in a New Zealand journal, could not necessarily be taken to indicate lack of quality.
- 110 The tie-point descriptors pointed to possible indicators of quality other than journal publication, but in practice peer review panels were not always able to deduce the quality of the RO from the evidence given. These problems in the use of proxies may explain the relatively poorer showing of practice-based research, New Zealand research, applied research, and performance and artistic research. Further study is required to test this inference and to judge the extent of the problem.
- 111 In Chapter 5 we make observations intended to assist any revision of the tie-point descriptors.

The scoring system

- 112 Peer review panels were first required to score each of the three components of an EP (RO, PE, CRE) separately on a 0-7 point scale. Tie-point descriptors were provided for the scores 2, 4 and 6. A score of 0 was awarded only if no evidence was supplied.
- 113 The scores for each category were then awarded a weighted score (with RO weighted at 70, PE at 15, and CRE at 15). A maximum score of 7 for each component translated into a maximum available score of 700.
- 114 The weighted scores were used to produce an Indicative Quality Category with boundaries set at 200 (R/C), 400 (C/B), and 600 (B/A).
- 115 Panels then made an holistic judgement to finalise the placement of EPs into Quality Categories using the following information:
- the three scores awarded;

- the Indicative Quality Category;
 - notes indicating uncommon factors contained in the EP; and
 - the overall information in the EP.
- 116 Certain rules constrained the final allocation of Quality Categories by peer review panels. These were:
- an RO score of at least 2 was required to award Quality Category C; and
 - an EP did not meet the minimum requirements for Quality Category C if the only NRO was a masters or doctoral thesis.
- 117 In practice, the scoring system produced some significant and unexpected distortions. The most significant of these are described in the following paragraphs.

New and beginning researchers

Recommendation 10

That the scoring system be amended to eliminate or mitigate the disadvantages and distortions affecting certain groups as described in this evaluation.

- 118 There is a broad consensus that new and emerging researchers were unfairly treated in the 2003 Quality Evaluation. This was largely due to an unanticipated effect of the scoring system, whereby it was difficult for young researchers to produce the evidence for the CRE and PE scores that would take them across the R/C threshold.
- 119 Despite many TEC clarifications of the 'R' Quality Category, the evaluation evidence is that it became almost universally understood as meaning 'research inactive'. This was particularly distressing for the many young researchers who were placed in that Quality Category. Early indications are that it has produced a morale problem amongst some staff in TEOs.
- 120 A Royal Society PBRF Forum (21 May 2004) reviewed this matter and proposed some solutions for consideration.

The significance of the PE and CRE components

- 121 As the work of the peer review panels progressed, it became clear that:
- the scores awarded for the PE and CRE components were more significant than had been anticipated at the boundaries between indicative Quality Categories; and

- many EPs had paid superficial, little, or no attention to the inclusion of evidence in the PE and CRE components.
- 122 Peer review panels were able to make holistic judgements, and were not bound by the indicative Quality Categories generated by the scoring (except as described above). However, in their holistic judgements they were required to use only the evidence contained in EPs. This meant that many EPs did not generate the Quality Category that may otherwise have been merited, because inadequate completion of the PE and CRE components resulted in lower scores; and these scores could not be corrected by an holistic judgement as this too had to be based on the information presented in the EP.

Streamlining the preparation of EPs

Recommendation 11

That the TEC:

- *develop an improved IT platform that will facilitate the creation, collection and processing of EPs, that will be effective across the range of platforms used by TEOs, that will minimise errors in data presentation, and that will minimise the costs of providing and maintaining PBRF data for TEOs; and*
 - *advise the sector of its intention to develop such a platform as soon as practical, so that TEOs can plan to integrate this platform into their internal systems.*
- 123 Now that TEOs have seen what EPs are used for, what they should contain, and how they are to be submitted, they are able to offer suggested improvements in the design and management of the EP preparation, assessment and submission process. TEOs expect to make substantial savings in their indirect costs from improving their own EP process and from integrating that process with their research recording, reporting and management systems.

The verification of NROs

Recommendation 12

That the suggestions by the National Library of New Zealand on the auditing of NROs be used as a basis for refining the EP compilation process.

- 124 The National Library of New Zealand provided the TEC with a report on its verification of NROs for the 2003 Quality Evaluation. We support the National Library's suggestions.



Preserving the PBRF Project Team's knowledge and competencies

- 125 A consistent message from TEOs was that the relationships established with members of the PBRF Project Team, and the support received from them, was a critical factor in carrying out the 2003 exercise successfully under the very tight time constraints. We found that there was an expectation amongst many in the sector that those relationships would continue, uninterrupted, up to the next Quality Evaluation.
- 126 We note that almost all of the key personnel who created these effective relationships were contract staff almost all of whom have been released. Sector confidence in the TEC's planning, operations and knowledge acquisition is at risk if the TEC loses momentum or if the sector concludes, early in 2005, that many of the lessons of 2003 are destined to be repeated. In our view, preserving sector confidence in the TEC could inhibit the development of oppositional thinking and strategies at the level of TEOs and individuals when some of the very hard issues, such as the appropriation of individual assessments for internal TEO performance management, need to be worked through.
- 127 We see merit in the TEC preserving relationships of the nature that TEOs enjoyed in 2003, at least until the end of the next Quality Evaluation and perhaps on a part-time basis. We have no recommendation on this, as the deployment of staff is the prerogative of the General Manager.

Chapter 1: Background to this evaluation

Introduction

128 This report is based on a complex case study that focused on generating evidence and analysis relating to the design and implementation of the PBRF in 2003 and the conduct of the 2003 Quality Evaluation. The Centre for Research on Work, Education and Business Ltd (WEB Research) undertook the case study and wrote this report.

Cabinet requirements

129 The Cabinet directed the MOE, after consultation with the TEC:

to report to the Social Development Committee by 31 March 2004 on the impacts of the PBRF as determined by modelling of impacts following the first Quality Evaluation, and to make any recommendations concerning:

- Changes to transition arrangements;
- The need for buffering of funding to providers;
- Investment required to maintain programmes of strategic importance;
- The impacts of the research completions measure on taught degree programmes and research quality, and any need for adjustment of the threshold for research components in the completions measure;
- The impact of the PBRF on the clinical education courses required to produce practitioners for the health sector; and
- Any other responses required (Tertiary Education Advisory Commission and the Ministry of Education, 2003, p4).

130 Despite a difficult schedule the PBRF assessment, including the 2003 Quality Evaluation, was conducted and completed within expectations. Publication of the results was delayed by legal action. For a variety of reasons, the MOE reporting time to Cabinet was extended to mid 2004.

131 The MOE and TEC agreed upon a high-level strategy document for evaluating the PBRF. *The Performance-Based Research Fund Evaluation Strategy Proposed by the Ministry of Education and the Tertiary Education Commission ('The Evaluation Strategy')* was issued in August 2003.



Evaluation Strategy proposed by the MOE and TEC

- 132 The MOE and TEC agreed to conduct an evaluation of the PBRF in three phases. The overall purposes of the three-phase Evaluation Strategy are to:
- determine the extent to which the aims of the PBRF have been achieved;
 - analyse the results of the Quality Evaluations (in 2003 and 2006) and assess what they reveal concerning the quality and pattern of research activity across New Zealand's tertiary education sector;
 - identify and assess the behavioural impacts, both positive and negative and both intended and unintended, of the PBRF since Cabinet approval (in December 2002), including the impacts on the nature, quality and quantity of research conducted in the tertiary education sector; and
 - provide evidence to inform policy decisions concerning the design, implementation and funding of the PBRF, including the transitional funding arrangements during 2004-2007, the conduct of the proposed Quality Evaluations in 2006 and 2012, and the PBRF funding formula (Tertiary Education Advisory Commission and the Ministry of Education, 2003).
- 133 This evaluation is part of Phase 1 of the planned Evaluation Strategy and this report contributes to Phase 2 of the Evaluation Strategy.
- 134 Phase 1 of the Evaluation Strategy is intended to focus on:
- an evaluation of the implementation process (especially in relation to the 2003 Quality Evaluation);
 - the short-term impacts of the PBRF on the tertiary education sector, including modelling the likely financial implications of the PBRF for TEOs during 2004-2007; and
 - the results of the Quality Evaluation and what these reveal about the overall quality of research being conducted in the tertiary education sector, the main areas of research strength and weakness, and the relative research performance of the TEOs that have participated in the PBRF.
- 135 This evaluation explored the evidence emerging for early indicators of:
- the success or otherwise of the 2003 Quality Evaluation framework;
 - whether the direct and indirect costs of the implementation of the PBRF and the 2003 Quality Evaluation were likely to exceed the

- minimum level consistent with a robust and credible assessment process;
 - the way in which TEOs responded and how this mediated the intent of the PBRF policy;
 - whether the 2003 assessment exercise was, and was seen to be, fair, reasonable and robust; and
 - the implications of the results.
- 136 The MOE will undertake the modelling of the likely financial implications of the PBRF for TEOs during 2004-2007.
- 137 In *The 2003 assessment* the TEC reported on the results of the Quality Evaluation and what those revealed about tertiary education sector research.

The role of the Evaluation Advisory Group

- 138 The Evaluation Advisory Group (EAG) is an essential component of the overall PBRF evaluation. Its role is to provide the independent evaluation team with professional expertise and advice from a range of stakeholder perspectives. The group is not, however, accountable for delivering the evaluation outputs within the resource limits set for Phase 1 of the Evaluation Strategy.
- 139 THE EAG met during the design phase of this evaluation to provide expertise and advice prior to agreeing on the evaluation design. EAG members were:
- Kaye Turner, Roger Staples (TEC)
 - Professor Jonathan Boston (Victoria University and TEC)
 - Barham Bekhradnia (Director, Higher Education Policy Institute, Oxford, UK)
 - Professor Michael Scriven (Professor of Education Evaluation, Auckland University)
 - Professor Deborah Willis (Professor Education and Assistant Vice-Chancellor [Academic] Victoria University)
 - Dr. Maureen Montgomery (Department of American Studies, Canterbury University)
 - Professor Mason Durie (Assistant Vice-Chancellor [Māori] Massey University and Chair of the Māori Knowledge and Development Panel)



- Elizabeth Eppel, Roger Smyth (Ministry of Education)
- Dr Roberta Hill, Phillip Capper, Ken Wilson (WEB Research).

140 The EAG met every six weeks during the evaluation. It also met to consider early drafts of this report.

WEB Research evaluation questions

141 WEB Research undertook an initial scoping exercise, in consultation with the EAG and stakeholders in the sector, in order to clarify the purpose of the evaluation, its focus, and the nature of the questions that could be addressed given the budget for this evaluation.

142 Trade-offs on evaluation priorities were required because of timing and sequencing constraints and other complex evaluation design issues that surfaced in the scoping stage.

143 In order to ensure the independence of the evaluation process, a judgement was made about research priorities. This included considering the relative weightings that should be given to implementation issues set out in the Evaluation Strategy and to emerging impacts and unintended consequences that the scoping exercise suggested would be needed both as data for Phase 2 of the Evaluation Strategy and as a means of meeting the concerns outlined in Appendix 2 of that report (Tertiary Education Advisory Commission and the Ministry of Education, 2003, pp13-19).

144 In summary, this evaluation focused on generating evidence and analysis that related to:

- **The appropriateness of the design of the PBRF framework:** whether the indicators of research quality were valid and reliable; and whether the Quality Categories were appropriate discriminators. Answers to this set of questions provide early indicators of short-term impacts and unintended consequences.
- **The conduct of the assessment of EPs:** whether the assessment of EPs was conducted in a robust, consistent and fair manner; if not, were the problems a consequence of the design of the PBRF, or its implementation?
- **The response of TEOs and individuals to the Quality Evaluation process:** how they responded and why; and whether and to what extent their responses were congruent with the overall intent of the PBRF.
- **The costs of the PBRF implementation:** what were the direct financial and indirect financial costs of the implementation of the

PBRF? (This work is to be carried out with the assistance of the Ministry of Education.)

- **The results of the 2003 Quality Evaluation:** what do the results reveal about the overall quality of research in New Zealand?

Undertaking this evaluation

145 Full details of the case-study-based design for this evaluation are given in Appendix 1. In summary, this evaluation was in two stages.

Stage One: Scoping

146 An initial scoping stage was needed for the development of the subsequent evaluation design. This decision was made in the context of the policy environment, the multiplicity and number of stakeholder interests, and the complexity of data sources. The scoping stage, based on interviews and the review of documents, clarified the objectives, research questions, evaluation constraints, and design for the second phase of this evaluation.

Stage Two: In-depth study of the issues

147 The in-depth study comprised:

- an analysis of direct and indirect costs;
- case studies of internal assessments in three TEOs;
- case studies of peer-review assessment processes (in four peer review panels, the two moderation panel meetings, and the two moderation review meetings), semi-structured interviews during the panel process, and a web-based on-line survey of all peer review panel members and chairs (83.3 per cent response rate);
- observation and understanding of results and early impacts (the delay in the public release of the results meant that our substantive fieldwork was completed before the results and impacts were evident – although we were able to observe very early responses and receive reports from TEOs on their strategic options); and
- synthesis of the evaluation data and production of the final report.

148 Those who are unfamiliar with case-study methodology are likely to be concerned at the apparently small number of case studies proposed for Components 3 and 4 in the list above. As Table 1 Appendix 1 shows, the overall case study is made up of seven components. Two of these components are, themselves, exploratory case studies of critical stages



of the PBRF implementation. When these case studies were combined with the other components, they produced a solid basis for Phase 1 recommendations or identified issues for further in-depth study in Phase 2 of the Evaluation Strategy.

- 149 The first stage of Phase 2 will involve an active engagement with TEOs, staff and stakeholder organisations around the issues raised in this report.

The nature and status of this evaluation

- 150 The evaluation evidence makes it clear that the full significance of design issues will emerge only after the PBRF has been fully implemented and the 2003 Quality Evaluation has been analysed. In the report, we have emphasised those issues that the evaluation data suggested are most likely to undermine the intent of the PBRF policy if they are not addressed.
- 151 The judgements and recommendations in this report are formed by triangulating the data, using analytic-inductive techniques as described in the methodology appendix (Appendix 1). Analysis of this data allows us to understand some of the key design and implementation issues and their possible implications. Other than in certain specified cases, however, we cannot draw inferences about the extent of the patterns we describe. That has to wait until subsequent phases of the evaluation of the PBRF.
- 152 Some of the evidence on aspects of the design and impacts is, however, indicative only. These aspects need to be evaluated further in Phase 2 of the Evaluation Strategy, while recognising that the PBRF assessment framework is complex and that some of the more profound impacts will emerge only in the medium to longer-term.
- 153 For WEB Research, the PBRF was understood to be a significant intervention in an existing, large, complex and diverse social system: that is, the system involving TEOs, government (agencies and politicians), and students. We had some prior knowledge of the existing tertiary system and we understood the nature of performance-based assessment tools in TEOs. But we did not have an understanding of the PBRF as a working model or as a policy tool.
- 154 We needed to understand the design of the PBRF and how it was intended to work at the very time that we were observing its impacts upon people and organisations. This provided us with a four-fold learning challenge. We were required to:
- understand what the elements of the PBRF were – that is, to learn the names, the differences that mattered (e.g. tie-points, Quality Category versus quality score, ERI), what the elements were



intended to do, and how all the elements were intended to work together;

- understand how the PBRF was intended to influence TEOs and academics in the medium and long term, in order to identify relevant early indicators of issues that would need to be resolved;
- observe the impacts of the implementation process and the results, and make judgements about how the TEOs initially viewed and responded to any such impacts; and then
- make a judgement as to the materiality and proportionality of our observations in the light of the policy and implementation intent.

155 This proved to be more difficult and to take longer than we could have anticipated. We began our work in September 2003 and needed to gather most of our data before the Christmas break. In retrospect, we would have benefited if this evaluation had commenced earlier in 2003 so that our grasp of detail, intent, impact and consequences would have helped us to identify the early indicators more efficiently.

156 One reason for our approach was that WEB Research undertakes its research, consultancy and evaluations using activity theory and developmental work research tools (Engestrom, 2000; Engestrom, Virkkunen, Helle, Pihlaja, & Poikela, 1996). We believed that our intervention in the implementation process should be developmental in its intent and not merely an instrumental or technical review of the PBRF as designed. We received considerable assistance from the EAG before the TEC and WEB Research agreed on the scope and nature of the activities to be conducted in an evaluation based on a developmental design.

157 We found that most people we interviewed in the TEOs had varying degrees of knowledge and that this was restricted to those aspects of the PBRF and Quality Evaluation they were involved in. Very few people had the detailed and holistic knowledge that could match the depth of knowledge held by key members of the PBRF Project Team. We needed to find a balance between the passion and clarity of the designers of the scheme and the frustrations and slight confusion of staff in TEOs who were seeking to read the manual and 'fly' at the same time.

Summary

158 In May 2002 Cabinet agreed to the establishment of the PBRF. Design work began and in December 2002 Cabinet endorsed *Investing in Excellence*. The implementation period then began. The PBRF Project Team undertook a comprehensive consultation process with the sector, to implement the PBRF scheme. Discussion papers led to decisions on

specific issues, to the issuing of the *Guidelines* and IT tools. The decisions included:

- definitions of who was eligible to participate in the quality evaluation;
- decisions on the number of multidisciplinary peer review panels and guidelines for selecting members;
- the evidence framework for the assessment, and the meaning of RO, PE and CRE;
- types of evidence that were acceptable, and definitions of the distinction between quality-assured and non-quality-assured evidence;
- a data-checking and verification process;
- responsibilities of peer review panel members;
- the assessment process to be used;
- the scoring system to be used and procedures for translating the scores into Quality Categories;
- a moderation process;
- a reporting framework; and
- a complaints process.

159 Given that:

- the task was complex;
- it was the first implementation of an entirely new process in New Zealand
- much was at stake for individuals and TEOs that heightened anxieties for both; and
- the timeline was tight,

it was inevitable that there would be elements which would not work as intended, that there would be unintended consequences, and that there would be some degree of inequity and unfairness for some participants.

160 Even without the novelty of the PBRF in New Zealand and the extraordinary timeline attempted, the complexity of the implementation of a new assessment scheme necessarily means resolution of the design over time. This has been the experience in other jurisdictions.



- 161 Given all that might have gone awry, it ought to be a matter of pride for all concerned that the PBRF was implemented in 2003 and the first Quality Evaluation was conducted simultaneously, with no major failure of the implementation process.
- 162 The evidence from interviews in TEOs, the TEC and other stakeholders suggest that the main contributing factor to the successful implementation of the PBRF and the conduct of the 2003 Quality Evaluation was the relationships established between many staff in TEOs and members of the PBRF Project Team. Some mutually supportive processes between administrative staff in different TEOs also contributed to a successful outcome.
- 163 The interview data suggests that those involved in the implementation process coalesced around a shared goal, a determination to 'get the job done'. The continuation of such a culture within the PBRF process would, in our view, support and facilitate a collaborative and developmental approach to addressing the more serious design issues that have emerged.
- 164 The continuing re-design of the PBRF and preparations for the 2006 Quality Evaluation could become an internal project with a technical focus for the TEC. Instead, we suggest that both Phase 2 of the Evaluation Strategy and the consultation process that follows this report should continue the TEC's established practice of enabling and supporting TEOs' contribution to the development of the PBRF.
- 165 In this report we adopted the view that an issue, even a serious issue, would need to present an immediate fatal risk to the PBRF to justify our recommending any fundamental design changes before the 2006 Quality Evaluation. Any proposed solution would need to demonstrate a benefit greater than the disruptive effect of changing a process that stakeholders had only one experience of. If it is too soon to say the PBRF is working, then it is also too soon to say it is not.
- 166 We have identified issues that pose a medium-term risk to the PBRF if left unaddressed, but we have not identified solutions that imply a fundamental re-design.
- 167 On the basis of the evidence generated in our overall case study and the understanding we have developed about the PBRF and the Quality Evaluation process, we have concluded that:
- the PBRF and Quality Evaluation should be retained for the next round;
 - the serious issues that have emerged should be recognised and addressed;



- continuous incremental improvements should be made to the scheme on the assumption that the scheme will not be significantly altered for the 2006 Quality Evaluation; and that
- the developmental process that this evaluation sought to engage the MOE, TEC and TEOs in should be continued in Phase 2 of the Evaluation Strategy.

Chapter 2: The individual unit of assessment and the reporting framework

Why the design decision was made

- 168 A key element of the PBRF assessment framework is the unit of assessment used for rating research quality. The Tertiary Education Advisory Committee and the PBRF Working Group considered three options for the unit of assessment during the design process: individual researchers; disciplinary-based groups or academic units; and the institution as a whole. The Tertiary Education Advisory Committee proposed that the unit of assessment should be the individual researcher.
- 169 This was recognised as a difficult option. In his report on the British and Hong Kong RAEs in 2003, Boston recommended that the TEC reconsider its decision to use the individual unit of assessment (Boston, 2002, pp29-35).
- 170 Boston reviewed the British and Hong Kong experiences and concluded that the advantages of using the individual unit of assessment were:
- it would be administratively simpler and involve substantially lower compliance costs;
 - it would be possible to operate an institutional form of 'self-assessment';
 - there were concerns about the capacity of multidisciplinary panels to assess the performance of disciplinary units;
 - there were concerns about the potential problems of rating disciplines in New Zealand's small tertiary system;
 - assessment of individuals fitted the framework of a mixed model of quality assessment (the proposed Quality Evaluation) and performance indicators (postgraduate degree completions and external research income); and
 - allocating funding on the basis of rating disciplinary groups was likely to lead to significant funding differences between similarly-sized departments in the same cost category but which had different ratings.
- 171 He reported the disadvantages as being:
- fairness to individuals demanded a higher degree of assessment rigour than in the case of groups;

- there could be negative impacts on collegiality and morale;
 - there was a risk that there would be appeals, including legal challenges, by aggrieved individuals;
 - there was a risk of class actions by groups that felt they had been collectively disadvantaged;
 - there was the potential for human resource management problems, especially individuals seeking to negotiate employment contracts on the basis of their Quality Category; and
 - the difficulty of rating individuals in disciplines where most research was done in teams.
- 172 The TEC confirmed that it preferred the individual unit of assessment. An advantage of the small New Zealand academic community, compared to much larger academic communities such as the UK, was that it might be possible to carry out an assessment of the quality of research at the individual level across the entire tertiary sector at an acceptable cost.
- 173 In this section we consider whether there are any issues associated with the individual unit of assessment that have emerged and what those issues might mean for the future of the PBRF.

The arguments for an individual unit of assessment

- 174 We begin by considering the experience of the 2003 Quality Evaluation against the design arguments for and against the individual unit of assessment. We then consider any other issues that emerged during the 2003 Quality Evaluation.

Was the individual unit of assessment administratively simpler and did it reduce compliance costs?

- 175 The UK RAE uses group assessment and the Hong Kong RAE uses individual assessment; both provide useful comparisons. However there are many variables that influence the costs in those RAEs other than the unit of assessment. The TEC noted that the administration and compliance costs of the 2003 Quality Evaluation appeared to be proportionately higher than those for the RAE and Hong Kong exercises;

In many respects the 2003 Quality Evaluation provided a much better way of assessing research quality than the Hong Kong RAE – but simplicity and low compliance costs were not among its virtues (Boston, 2004, p6).

- 176 The sector and the TEC have now developed some expertise in working with an individual unit of assessment and developed some systems for handling individual EPs. The 2003 Quality Evaluation could be regarded as a research and development exercise, especially around identifying improvements to the completion and management of EPs. The time and the costs invested by TEOs represent a sunk cost that may be lost if the TEC moved to a group assessment model.
- 177 At the Royal Society of New Zealand PBRF Forum (RSNZ Forum, May 2004) Boston suggested that the New Zealand model has proved superior to many others (Boston, 2004). International members of peer review panels interviewed in this evaluation praised the design and method of the individual assessment and noted its advantages over their own experiences.
- 178 If the individual unit of assessment proves to be a better system for assessing research quality for New Zealand than other units of assessment, then whether the costs of using that system reach an acceptable level will remain open until the costs are more settled, perhaps following the 2006 Quality Evaluation.
- 179 The administrative and compliance costs of the implementation of the PBRF and the 2003 Quality Evaluation were a very high percentage of the funding allocated TEOs by the 2003 Quality Evaluation for the period 2004 to 2006. The percentage of PBRF funding that costs make up will diminish rapidly as the amount of PBRF funding rises each year through to 2007.
- 180 The analysis in Chapter 4 indicates that administrative and compliance costs of the 2006 Quality Evaluation and the ongoing costs of the PBRF to the MOE, the TEC and TEOs in the period 2007 to 2012, are likely to approach a level that is comparable with the administrative and compliance costs of the UK RAE.

Has using the individual as the unit of assessment enabled there to be an element of self-assessment?

- 181 At the Royal Society PBRF Forum, Boston observed that enabling a process of self-assessment in TEOs has emerged as unsustainable.

With hindsight, the whole idea of self-assessment was rather ambitious, if not misguided. On the basis of the experience and results of the 2003 Quality Evaluation, an assessment model based largely on internal TEO evaluations of their staff would be hard to implement in an effective, consistent and credible fashion. Moreover, I have little doubt that we would need a comprehensive and exacting system of external audit, and that to operate such a system effectively would probably

require a panel structure and process not unlike the current model (Boston, 2004, p7).

- 182 The comparison of the Quality Categories assigned by TEOs in their internal assessment with the final Quality Categories assigned by the peer review panels demonstrates the difficulties that lie ahead of the internal self assessment process.
- 183 Internal peer reviews were intended to assist TEOs and staff to understand and apply the assessment processes; they were in part, a learning strategy. Internal peer review panels were seen as continuing after the peer review panels were disbanded. This plan was to reduce the costs of the individual unit of assessment by removing the expensive peer review panels. Experience now indicates that the peer review panels will need to be retained and that it is the internal peer review panels that may cease. This may bring some savings in TEO compliance costs and staff time if TEOs do not elect to retain internal panels for their own purposes.

Could multidisciplinary peer review panels assess the performance of disciplinary units?

- 184 This is a proposal that has not been tested in practice. Boston reports that TEAC originally argued that multidisciplinary panels were unavoidable in New Zealand's small system, and that it was then felt that 'the assessment of academic units under a multidisciplinary panel structure would be more problematic than the assessment of individual researchers' (Boston, 2004, p7).
- 185 The relationship between multidisciplinary panels, the unit of assessment and the nature of the assessment framework are complex interrelated matters. Drawing on international models of assessment processes provides some assistance to the technical design of an assessment model for New Zealand. There are strong situational and cultural dimensions in New Zealand workplaces, including universities, which will need to be recognised and considered if any change in the unit of assessment is considered.
- 186 This evaluation shows that the assessment model has been well-supported in practice and strongly supported by peer review panel members (On-line Survey, Appendix 2).

Does the assessment of individuals better fit the mixed model of peer review and indicators than a group-based approach?

- 187 This question is a matter of design philosophy. If academic units are the unit of assessment, then multiple indicators of quality are logically best applied at the level of each unit. If research degree completions

and external research income are considered at the institutional level, then the individual unit of assessment of peer assessment is more logical.

Has the funding allocation produced equitable outcomes?

- 188 The timing and limits to this evaluation mean we have no data to enable us to comment on this matter.

The arguments against the individual unit of assessment

Has the need to be fair to individuals produced an unsustainable or excessive requirement for rigour in the process?

- 189 Whether the assessment process is unsustainable in terms of its administrative, compliance and staff time costs is discussed in Chapter 4. Whether the need to be fair to individuals produced an unsustainable or excessive requirement for rigour in the process will remain difficult to assess until after a second Quality Evaluation provides a basis for comparison.
- 190 Some problems with fairness emerged in the practice of the assessment process. Those fairness problems were associated with the interpretation and application of definitions, the *Guidelines*, and the scoring frameworks, they were not reported as evidence of an excessive requirement for rigour.
- 191 The audit processes were reported as rigorous in the case of NROs but less so for the staff eligibility audit. The MOE and the TEC have indicated that the audit processes and instruments are to be reviewed with a view to improving their effectiveness and reducing the compliance costs to TEOs.

Have there been negative effects on collegiality and morale?

- 192 In submissions to us after the release of the 2003 Quality Evaluation results there were indications of impacts upon morale but few reports of impacts upon collegiality. Whether these reported impacts are an inescapable consequence of using the individual as the unit of assessment or are impacts that are remediable through adjustments to the definitions, descriptors and the *Guidelines* used to assist the quality assessment, is a judgement we could not make. We noted that the UK RAE is considering the more profound question as to whether collegiality and morale problems might be associated with Quality Evaluations *per se*, irrespective of the unit of assessment.



- 193 Whether any TEOs or individuals alter their behaviours and strategies as a consequence of the implementation of the PBRF and in order to benefit from the periodic Quality Evaluations are matters for Phases 2 and 3 of the Evaluation Strategy; it is simply too soon for this evaluation.

Have there been challenges and appeals by aggrieved individuals?

- 194 There have been 41 complaints about the Quality Categories awarded in the 2003 Quality Evaluation. However the criteria for lodging a complaint were that:

The TEC will accept and investigate only those complaints that concern possible administrative or procedural errors' (Tertiary Education Commission, 2004a, p269).

- 195 The narrowness of the grounds for appeal has been raised as an issue of fairness with us. Improvements to the assessment process and our suggestions, later in this section, for a Code of Practice, are intended to reduce the stakes for individuals of using the individual unit of assessment. The proposals are intended to reduce the risk of the misuse of the results and especially any formal or informal penalties on staff, i.e. lower academic status, less secure tenure, slower promotion, lower salary or fewer research opportunities, because of the misuse of the individual unit of assessment.

Are there any indications of class actions by groups that feel they have been disadvantaged?

- 196 No participant raised the possibility of a class action during the evaluation. This may not mean such actions are not being considered, rather it may reflect the timing of our fieldwork in TEOs, the later publication of results and the time it will take for groups to identify systemic disadvantages to base such a claim on.

Are there any indications that individuals are seeking to renegotiate their contracts of employment on the basis of their Quality Category?

- 197 No employee in the cases studied reported that they were seeking to renegotiate their employment contract on the basis of their assigned Quality Category. The results had not been published at the time of our fieldwork in TEOs.
- 198 Some employers in the cases studied expected some employees to use their assigned Quality Category in future negotiations and were not sure how they would respond.

Did it prove difficult to rate researchers in subjects where most research is done in teams?

- 199 Some participants in the evaluation raised concerns that in subjects where most research is done in teams, the panels would not be able to rate the research fairly or accurately. We observed that this matter was canvassed in peer review panels in respect of some EPs. The matter of rating research done in teams is discussed in more detail in Chapter 5.

Unintended consequences

- 200 During this evaluation we considered a range of evidence that suggests that individual participants in the PBRF 2003 Quality Evaluation, the management and administrative staff in TEOs and the public particularly, understood an individual quality score to be a personal performance grade.
- 201 Indicators from the operation of peer review panels, from the TEC reporting of results, from public interest and comment (partly led by media analysis), from TEO human resource planners, managers and from individual researchers themselves suggests that the individual assessment is highly likely to be taken out of context and out of proportion. Individual results are very likely to be misappropriated, formally and informally, for purposes other than those intended or required in the overall PBRF scheme.
- 202 By 'out of proportion' we are referring to the phenomenon highlighted by a New Zealand senior academic in one of our case study interviews: 'The PBRF is a measure of quality, not the measure'. This echoes the point made by a senior manager of a TEO that the criteria used in the 2003 Quality Evaluation 'are only some of the means by which research quality may be assessed in a changing institutional environment'.
- 203 The newness of the PBRF, its title, 'performance-based' which resonates so strongly with common sense notions that the good should be rewarded, the reporting using the letters A-R which echoes everyone's experiences of schooling and the utility of an independent individual assessment all make it very attractive to use the Quality Categories for purposes for which they are not intended.
- 204 The perils of performance or merit based schemes that are intended to be measures of individual performance are well known (Murnane & Cohen, 1986). The Quality Evaluation is not intended to be a measure of individual performance for performance management and there will be a very real struggle to ensure that it remains a tool used only for its purpose.
- 205 Some individuals interviewed noted that they could support the use of the individual unit of assessment for PBRF purposes. They noted that

they would not support the PBRF in the future if managers within TEOs used the results of that assessment to manage staff.

- 206 There is close relationship between the teaching and research duties allocated to a staff member and their academic achievement, their productivity and their promotion in their TEO. It was clear from our scoping interviews, case studies and submissions, that some staff in universities consider the PBRF may provide managers with more power and control over their careers than is the current case.
- 207 Examples of the misappropriation of the purpose of the individual unit of assessment and its result have emerged. These include:
- sensational and shallow analysis and reporting of the results of the 2003 Quality Evaluation by the media;
 - national advertising by some TEOs making claims about their relative performance based on a partial use of the published results of the 2003 Quality Evaluation and other elements of the PBRF (RDC, ERI). This was best illustrated by the wry observation to us that an unexpected benefit of the PBRF reporting framework in 2004 was that as every university could show that it was the best in something, no university lost out;
 - inadvertent potential exposure of individual participant scores to any interested observers in a number of situations. Examples include all members of an academic unit getting the same score; all but one member of an academic unit getting the same score and the exception voluntarily revealing his or her score; and academic units of only one person; and
 - recognition that the small number of requests by individuals for their personal scores from the TEC, including their initial rating by the pairs, could become an industrial tactic for a much larger number of staff in different circumstances.

The reporting framework

- 208 The reporting framework is connected to the question of the choice of the individual unit of assessment. Although the TEC expected difficulties in developing a satisfactory reporting framework, Boston (Boston, 2004) has reported that these problems were of a greater magnitude than had been expected.
- 209 *The 2003 Assessment* describes and addresses these difficulties (Tertiary Education Commission, 2004a, pp36-37).
- 210 The results of the 2003 Quality Evaluation were reported at the four levels that had been indicated in the *Guidelines*. These levels were TEO,

panel, subject area and nominated academic unit. The issue of concern was the discovery that it would be easier than had been anticipated for an interested observer to infer individual Quality Categories from the published results and using complementary information obtained from other sources.

- 211 There have been other difficulties associated with the reporting framework. These include:
- the question of the validity of the 41 subject area groupings chosen;
 - the inconsistency between TEOs in the choices made about how to group staff into academic units; and
 - the inability to report demographic data because not all TEOs provided the necessary information.
- 212 When it recognised that individual Quality Categories might readily be inferred, the TEC considered a number of ways of reporting the results to reduce that risk of exposure for individuals. The options considered are reported in *The 2003 Assessment* (paras 104-106). The TEC finally resolved to report as it had intended because no other option was satisfactory.

Summary

- 213 From an evaluation point of view the questions are:
- have the fears that widespread inference of individual scores would be possible been justified?
 - If so, what risks does this create for the future success of the PBRF?
- 214 There has not been time to measure systematically the extent to which widespread knowledge of individual Quality Categories has occurred; it would be difficult to do so accurately as well. Anecdotal information suggests that it has happened:
- where all members of an academic unit have received the same Quality Category;
 - where all members but one in an academic unit receive the same Quality Category, and the exception reveals his or her own category; and
 - where academic units are very small.
- 215 The risks associated with widespread – or even limited – disclosure or inference of individual Quality Categories are:



- negative impacts on collegiality and morale;
 - significant levels of complaint, appeal, or recourse to the legal system to challenge results;
 - misuse of knowledge of individual Quality Categories by employers and employees;
 - all of which almost certainly lead to loss of trust in and support for the PBRF in the sector.
- 216 This evaluation indicates:
- that the problem of inference and disclosure was reported, but we can not report on the extent of that;
 - that the consequential risks exist, but it is too soon to establish the extent to which they are occurring in practice.
- 217 There is insufficient evidence to determine whether a case exists for major redesign of the unit of analysis.
- 218 However the misuse, intentional or inadvertent, of the results of using the individual as the unit of assessment is a significant threat to the success of the PBRF. Individuals in New Zealand TEOs operate within legal (industrial, professional, personal) frameworks that provide them with the means to oppose, frustrate and over time, weaken the PBRF if they chose to do so.
- 219 Short of moving away from using the individual as the unit of assessment, a major re-design effort, we suggest two actions for consideration by the TEC.
- There are a number of modifications to the assessment and reporting frameworks that if taken, may mitigate the risks detailed above. These modifications are contained in recommendations and proposals in this report and would form part of the various matters to be included in the TEC's review.
 - The TEC could sponsor the development of acceptable use of results codes of practice to apply to the TEC, TEOs and individuals. Such codes would have moral rather than legal force, but the moral force would be substantial if such codes were developed collaboratively and supported by all stakeholders.

Chapter 3: Eligibility of individuals and TEOs

Staff participation - criteria and definitions

- 220 Determining the number of staff eligible to participate in the 2003 Quality Evaluation was a fundamental element of the design and implementation of the PBRF in 2003. *Investing in Excellence* discussed the planned scope of the participation criteria for providers and individuals (Ministry of Education and Transition Tertiary Education Commission, 2002, pp11-12). The TEC then determined the appropriate criteria and issued those criteria to the sector, and TEOs exercised their judgement as to which staff were eligible.
- 221 The Staff Participation Criteria (eligibility criteria), including the substantiveness test (defined below), were detailed in Strategic Consultation Draft Guidelines posted to the TEC web site on 31 May 2003. The proposed criteria had been circulated to the sector for consultation in an earlier paper.
- 222 The Draft Guidelines for the 2003 Quality Evaluation were consolidated and issued to TEOs on 25 July 2003 as the '*Performance-Based Research Fund: A Guide for 2003*' (the *Guidelines*) (Tertiary Education Commission, 2003). The *Guidelines* did not alter the eligibility criteria but inserted the substantiveness test in the Staff Participation Criteria Process Diagram.
- 223 For TEOs in the 2003 Quality Evaluation, identifying PBRF-eligible staff was:
- a compliance task. TEOs were required to identify PBRF-eligible staff in accordance with the *Guidelines* and obtain EPs from those staff; and
 - a strategic task. TEOs were in a position to influence their overall Quality Score by the way they chose to interpret and apply the criteria in the *Guidelines* and particularly how they applied the substantiveness test.
- 224 Determining the total number of PBRF-eligible staff is critical for a TEO because all PBRF-eligible staff are included in the denominator for calculating the quality score. However, under the current policy framework a TEO faces conflicting pressures in applying the eligibility guidelines. On the one hand, it has an incentive to minimize the number of eligible staff – especially staff likely to secure an 'R' – in the interests of reducing the size of its denominator and thus increasing its likely quality score. On the other hand, if it excludes staff members that have a chance of securing at least a 'C', it runs the risk of reducing the funding to which it is entitled.

- 225 However, as we have discovered during this evaluation, the workings of the PBRF are complex and subtle. The denominator does not work as simply or in the simple causal manner as the preceding paragraph suggests. Funding for the quality measure of the PBRF is allocated on basis of the number of As, Bs and Cs awarded to each TEO (weighted by the relevant staff member's FT and subject area). The total number of PBRF-eligible staff in a TEO is used only as a denominator to determine each TEO's average Quality Score (ranking) and is unrelated to the level of funding allocated under the PBRF mechanism.
- 226 It appears that TEOs may have adopted different strategies in their approach to the application of the eligibility criteria for the 2003 Quality Evaluation. Some of those different strategies were informed by early and partial understandings of the eligibility criteria and the dynamic between PBRF-eligible staff numbers and the whole funding scheme. TEOs will be better informed and prepared for the 2006 Quality Evaluation, and thus may adopt different strategies.

Staff participation – data checking and verification

- 227 A framework for Data Checking and Verification was set out in Part 8 of the *Guidelines*. That framework would be 'focussed on the types of data where inaccuracies pose(d) the greatest risk to the integrity of the PBRF. Those areas include(d) staff eligibility to participate in the PBRF.'
- 228 The Validation and Verification of Staff Eligibility (eligibility audit) was a responsibility of the MOE and involved three stages:
- MOE validated the data in the PBRF Census (staffing return) and then forwarded the data to the TEC;
 - the TEC validated the data in the EPs that it received against the PBRF Census; and
 - the MOE undertook an audit of staff eligibility, including the application of the 'substantiveness' test by TEOs on behalf of the TEC.
- 229 The *Guidelines* noted that the MOE eligibility audit might include:
- random checks and requests for TEOs to supply extracts from their payroll and personnel records, including staff employment agreements; and
 - requests for TEOs to justify their inclusion, or exclusion, of certain staff.
- 230 The *Guidelines* indicated that wherever significant errors or discrepancies were detected, additional data checking would be

undertaken. The MOE might also compare the proportion of research fellows and teaching fellows that had been deemed to be eligible for inclusion in the PBRF across TEOs. Major disparities would be investigated. Provision was made for the application of sanctions.

- 231 The provision for validation and verification of staff eligibility was a critical element of the PBRF for TEOs. TEOs in our evaluation expected the eligibility audit to offer:
- independent confirmation of their interpretation of key definitions around research and teaching in the substantiveness test; and
 - independent confirmation of the interpretation of key definitions around research and teaching in the substantiveness test applied by *other* TEOs.
- 232 The eligibility audit was a mechanism to provide:
- assurance to the TEC that the appropriate staff had been included in the 2003 Quality Evaluation; and
 - assurance to TEOs that all TEOs were entering the 2003 Quality Evaluation on an equal basis especially as the eligibility criteria were new, open to interpretation and perceived to be capable of being manipulated.

Implementation of the Guidelines for determining PBRF staff eligibility

- 233 TEOs conducted their internal assessment of EPs in the period between May and September 2003. The PBRF-eligible staff were determined by applying the eligibility criteria set out in the 31 May draft guidelines since the final version of the *Guidelines* was not available in time.
- 234 TEOs in our evaluation reported that they needed to clarify the eligibility of their staff well before completing the 31 July 2003 *PBRF Census: Staffing Return* (PBRF Census). Since the eligibility audit was expected to begin in August 2003, immediately after the Census, TEOs were concerned to ensure they had complied with eligibility criteria and applied the substantiveness test accurately before the PBRF Census.
- 235 Some TEOs reported that:
- they found the initial Guidelines (i.e. 31 May 2003) difficult to understand and translate into simple instructions for staff;
 - they found the key definitions around research and teaching in the substantiveness test ambiguous; and

- they were not able to secure adequate clarification from the TEC about the application of the eligibility criteria to staff at the boundaries of the substantiveness test. This caused those TEOs to become concerned about the interpretation of eligibility being applied by other TEOs.
- 236 Some TEOs reported that they expected that the revised (25 July 2003) Guidelines would resolve their questions about eligibility. When this was not the case, the TEOs reported that they were concerned about:
- how confident they could be that their own application of the eligibility criteria was accurate and consistent and would not, when audited, result in censure or penalty; and
 - their recognition that if they were unable to clarify the eligibility criteria for themselves, then there was a possibility that the criteria might be applied inconsistently by other participating TEOs. Some TEOs believed that the inconsistency could result in the final quality scores and funding allocations being unfair and inaccurate.
- 237 The warrant against the fear of such 'gaming' was the eligibility audit to be conducted by the MOE on behalf of the TEC. TEOs expected the audit to be searching and rigorous and to result in sector wide confidence that all TEOs were applying the eligibility criteria consistently.
- 238 Following the publication of the results of the 2003 Quality Evaluation in April 2004, a number of TEOs made submissions to us that expressed reservations about whether the eligibility criteria had been applied consistently by participating TEOs.
- 239 In May 2004, the TEC wrote to TEOs addressing *Concerns over the Integrity of Staff Eligibility Data and the Staff Eligibility Audit*. The letter was a formal response to concerns about eligibility raised by the sector. The letter addressed a range of concerns including rumours and analyses of participation rates at various TEOs known to be circulating. TEOs with specific concerns regarding the nature or application of the PBRF staff eligibility guidelines were invited to draw those to the attention of the TEC.
- 240 The TEC letter noted that:
- comparisons of data in the MOE's single data return (SDR) and the data on staff participation in the TEC's report on the results of the PBRF were not reliable because the definitions and criteria used in the SDR are different from the definitions used to determine PBRF eligibility;
 - the MOE is giving serious consideration to changing the SDR definitions to bring them into line with the requirements of the PBRF;



- the TEC had confidence in the integrity of the audit conducted by the MOE;
- all TEOs co-operated fully and helpfully with the conduct of the audit and all issues and concerns that were brought to the attention of the audit team were investigated and satisfactorily resolved;
- no evidence was found, nor brought to the TEC's attention, which suggested that there had been systematic or widespread non-compliance with the staff eligibility guidelines;
- the TEC, in consultation with the MOE, had decided to conduct a relatively non-intrusive staff eligibility audit to minimise costs, to reduce demands on TEO staff under pressure from the implementation timeline and because the audit of NROs was relatively demanding;
- there is a distinction between non-compliance with the PBRF staff eligibility guidelines and actions by TEOs that were designed to enhance a TEO's likely quality score but that were consistent with the *Guidelines*. The TEC intends to review this matter in Phase 2 of its evaluation of the PBRF and the 2003 Quality Evaluation; and
- the TEC assured TEOs that it intended to review the substantiveness test prior to the 2006 Quality Evaluation.

241 The TEC advised us that, thus far, no formal submission or complaint has been made to the TEC about the application or operation of the staff eligibility criteria in 2003.

Staff Participation Criteria – Staff must be an employee

242 The criteria for staff participation are set out in Part 2, Section B of the *Guidelines*. The first condition is that eligible staff must be employees of the TEO. The criteria for employment provide that:

staff will be eligible to participate in the Quality Evaluation process if they meet all of the following criteria. They are:

- a **EITHER** employed on the staff census date by a TEO or eligible subsidiary under an agreement of salaried employment with a duration of at least one year
- OR** have been employed on the staff census date by a TEO or eligible subsidiary for at least one year under one or more agreement(s) of salaried employment on a continuous basis

AND

- b they are employed for a minimum of one day a week on average or 0.2 FTE over the period of the entire year

AND

- c their employment functions include research and/or teaching degree-level programmes
- 243 These three employment criteria are followed by definitions of some of the terms contained in those criteria. The terms further defined are:
- o Employment agreement requirements;
 - o Exemption for staff permanently located overseas;
 - o Continuous employment;
 - o Employment functions; and
 - o FTE status.
- 244 TEOs reported some concerns with:
- o the definition of continuous employment. One concern was that it would be a simple matter for a TEO to ensure that some categories of staff who they wished to exclude from eligibility, such as research assistants and tutors, had a break in service exceeding one month and were thereby ineligible for the PBRF.
 - o the definition of employment functions. There were concerns about the internal processes TEOs use to set the employment functions of staff and whether those result in accurate descriptions of the actual work staff undertake; and
 - o the setting and calculation of the current 0.2 FTE level. This concern pointed to the larger question about the intended scope of the eligibility criteria discussed below.
- 245 TEOs saw these concerns as matters able to be resolved before the 2006 Quality Evaluation in the consultative process proposed by the TEC. Some staff considered that how employment functions might be set in a TEO was a matter for possible consideration in industrial bargaining because the advent of the PBRF had raised the significance of agreeing to those functions.
- 246 The application of the eligibility criteria in the 2003 Quality Evaluation by TEOs showed that some hold different views about which staff were intended to be included or excluded from a Quality Evaluation.
- 247 Proportionately fewer staff suggested that the substantiveness test should be reviewed in order to include more of the staff who were excluded from the 2003 Quality Evaluation. Some TEOs suggested that the substantiveness test should be reviewed in order to exclude more of the staff who were included in the 2003 Quality Evaluation.

- 248 The participants proposed both views as mechanisms for improving the PBRF and reducing costs. These quite different views of the eligibility criteria illustrate the need for a more accurate common understanding in the sector about:
- the intended scope of the eligibility criteria;
 - the operation of the eligibility criteria; and
 - the scope of TEO discretion in applying the eligibility criteria.
- 249 Reducing the number of eligible staff was argued to be a proposal that might:
- reduce the costs of the PBRF for TEOs and the TEC by decreasing the number of EPs to be prepared and assessed;
 - spare TEOs and staff from investing time and energy in avoidance strategies and behaviours that may be inimical with good HR practices; and that might
 - avoid awarding 'Rs' to some categories of staff who appeared to have been included in the 2003 Quality Evaluation for no sensible reason.
- 250 The tight implementation timeline in 2003 meant that TEOs had to apply their understanding of the eligibility criteria as best they could. As there was no pilot and the stakes of decisions around eligibility were perceived to be high, there was some anxiety in some TEOs about decisions they had to make for staff at the margins of the eligibility criteria.
- 251 The sector expects to discuss and resolve issues around the employment criteria with the TEC before the guidelines for the 2006 Quality Evaluation are finalised.

The definition of an employee

- 252 TEOs have a variety of service delivery structures that influence employment arrangements. TEOs may own subsidiaries with various degrees of ownership and they may second or share staff with entities that are not within their span of control. One of the design elements of the PBRF that distinguishes it from the UK RAE is that in New Zealand it is the activity – research - that determines eligibility, not the nature or source of the funding of that activity. Defining a researcher as an employee is a method for capturing the activity irrespective of the funding. However this has brought some challenges in reconciling a legal state of being for an individual (I am an employee) with a cultural state of being for the same individual (I am an academic defined by my research/teaching).



- 253 There have been major changes to the employment framework in New Zealand following the Employment Contracts Act (1991) and the Employment Relations Act (2000). These Acts permit and enable a variety of forms of employment relationships that TEOs reported were not captured by the definition of employment in the eligibility criteria. Staff who are not employed on a permanent full-time basis may be employed as casual, casual part-time, permanent part-time and even permanent casual part-time.
- 254 TEOs reported that the employment status of some staff, such as research fellows, research assistants and professional staff on secondment to or from another organisation, was difficult to resolve. One example of uncertainty was research fellows who may have a major research responsibility in a TEO but are not paid by the TEO. Applying the eligibility criteria means they are not employees and should not be included in the Quality Evaluation. In these circumstances the employment criteria apparently exclude some 'high' quality researchers from the PBRF.
- 255 We were advised that some TEOs had reviewed the employment status of certain staff prior to the date of the PBRF Census. The reported effect was that some staff who had earlier been considered to be eligible for the PBRF were now considered to be ineligible.
- 256 Within the limits of this evaluation we observed no examples of creative or mischievous interpretations of the eligibility criteria.
- 257 The operation of the eligibility criteria provided TEOs with the responsibility of exercising their discretion in the application of the employment criteria and the substantiveness test. Logically, TEOs were best placed to make decisions about eligibility and especially the substantiveness of an individual's contribution, and that will remain the case.
- 258 The observations in this evaluation show that some TEOs were hesitant about taking up that responsibility.
- 259 Some TEOs in this evaluation indicated that they intended to review their employment arrangements and the terms and conditions of employment of staff in the future in order to comply with the employment criteria and to reduce the number of PBRF-eligible staff. This evaluation could not determine if this was to be staff who are not expected to secure at least a 'C' or a more general strategy.
- 260 Whether TEOs do alter their employment practices and whether that alteration has a substantial impact upon the overall Quality Score of a TEO or the careers of researchers is a matter for monitoring and review in Phase 2 of the TEC's Evaluation Strategy.



- 261 The speed of the implementation process and the unfamiliarity of TEOs with the staff participation criteria did not allow for informed fine-tuning of the eligibility criteria during the 2003 Quality Evaluation. There was no time to ensure that the criteria were sensible or that they were properly and consistently applied.
- 262 There remain complexities around the range of employment arrangements used in TEOs that require review. TEOs, MOE and the TEC all have the benefit of a grounded understanding of the operation of the PBRF's current eligibility criteria. We support the TEC's declared intention to review the intended scope and application of the employment and eligibility criteria before the 2006 Quality Evaluation.

The substantiveness test

- 263 Criterion C of the employment criteria set out in the *Guidelines*, deems employees whose 'employment functions include research and/or teaching degree-level programmes' to be eligible for inclusion in the PBRF.
- 264 The employment criteria were a first sieve for eligibility: all PBRF-eligible staff who met all of the three criteria were deemed to be PBRF-eligible. However, a person who met criterion a) – employed on Census date or have been employed on Census date, and b) – employed for one day a week or 0.2 per annum, was also required by criterion c) - to *include* research and/or teaching degree level programmes in their employment functions.
- 265 Because the set of employees in a TEO whose employment functions might include research and/or teaching degree level programmes could be very large indeed, the *Guidelines* added a second sieve for eligibility, a test of the substantiveness of that research and/or teaching contribution. The object of the test was to establish that an employee must make a substantive contribution to degree-level research and/or teaching; a small or insignificant contribution would not be sufficient to make an employee, PBRF-eligible.
- 266 The *Guidelines* detail how TEOs were required to apply the substantiveness test to senior administrative staff, staff who taught both degree level and sub-degree level concurrently, tutors, laboratory assistants, technicians and other technical support staff and research assistants.
- 267 The specification of roles was expected to address the eligibility of most PBRF relevant roles currently found in a TEO. The eligibility of staff in roles that were not specified, and no list could specify them all, was for the TEO to determine using the substantiveness test.

268 TEOs read and interpreted the substantiveness test as an authoritative expansion of the meaning of criterion c) of the employment criteria. The substantiveness test was to answer the question, what does 'research and/or teaching degree-level programmes' mean?

269 The substantiveness test states:

A staff member who fills one of the job categories above should be included in the Quality Evaluation process if, in the 12 months preceding the census date the staff member is:

- Undertaking teaching at degree level that includes (our emphasis) the design/updating and/or major teaching role and/or assessment of courses

and/or

- Undertaking research as a requirement of his/her employment that includes (our emphasis) the design or conduct of research activity and/or preparing research outputs (e.g. as a co-author/co-producer)

Note: Any research tasks considered under this test must conform to the PBRF definition of research.(see page 16 of this Guide).

A staff member who fills one of the job categories above would be excluded from the Quality Evaluation process if she or he is working under strict supervision of another staff member while teaching (e.g. working only with small groups of students in tutorial sessions or marking papers to strict criteria) or is involved in activity supporting research where there is no direct contribution to the design or conduct of research activity and/or preparing research outputs (e.g. providing laboratory services).

Research and teaching in the substantiveness test

270 The substantiveness test emerged as an issue during the 2003 Quality Evaluation for two reasons:

- some TEOs and managers remained uncertain as to what it meant and how they should apply it. This caused some TEOs and managers some anxiety at a critical time in the 2003 Quality Evaluation. The reasons for their uncertainty are considered in this section. Most of their concerns about the substantiveness test are likely to be resolved in the review proposed by the TEC; and
- some TEOs believed that other TEOs took advantage of the uncertainty of the application of the substantiveness test and the light-handed audit of staff eligibility, to increase their share of the PBRF funding allocated. There is work to be done on establishing a platform of high trust in the sector around the consistent application of the eligibility criteria that makes the fairness of that application visible.

- 271 During the evaluation we were advised that many TEOs believed that they had applied the substantiveness test consistently with the *Guidelines* and that they had not found that difficult. We were also advised that some TEOs found the test ambiguous and that they had not been able to apply it as precisely as they wished.
- 272 TEOs approached the task of interpreting the substantiveness test from their perspective, that perspective was strategically and culturally determined in some instances. The test was a technical solution to a difficult but critical issue, it was also an attempt to express a test of substantiality without resorting to specifying proportions. For example what constituted teaching/research was not defined as teaching 2 of 4 semesters, or 13.75 teaching weeks per annum and so on unto despair.
- 273 The substantiveness test was intended to clarify eligibility by narrowing what counted as research and/or teaching at degree-level programmes and exclude employees who made an insubstantial contribution, but it was understood to be intended to have the opposite effect by some TEOs.
- 274 Some TEOs read the PBRF definition of research as new and as extending the set of activities that would count as research for the purposes of the PBRF. Those TEOs approached the substantiveness test with an expectation that what they understood to be research was the activity that the test (definition) referred to when it defined research and/or teaching.
- 275 Thus TEOs with large numbers of professional staff, for example colleges of education, the applied schools such as physiotherapy and medicine, wānanga and polytechnics, regarded staff as eligible because of the respective TEO's interpretation of what counts as research. Given that interpretation and the need to comply with the *Guidelines*, those TEOs tended to apply the substantiveness test so that more staff were included than may have been required.
- 276 Some managers noted that the teaching and research activities in the test are all alternatives because they understood and/or to mean one or the other, not both. Thus a staff member need only undertake one of the listed alternatives to be PBRF-eligible.
- 277 Some managers read the test to say that a staff member was PBRF-eligible if their role 'included' any of the specified activities. As the test does not provide a measure of how much of any specified activity is substantial, they understood the test to say that if your role included any of the specified activities then that was the test and definition of substantiveness.
- 278 This interpretation would turn the substantiveness upon its head. Instead of the test being understood to specify and set limits upon what

counted as a substantial contribution to research and/or teaching, the test would be understood to deem any of the research and/or teaching activities specified in the test that were part of a staff member's role as making them PBRF-eligible.

- 279 These interpretations caused some TEOs to believe that a very wide range of their employees was now PBRF-eligible by virtue of the substantiveness test. It was not surprising that TEOs, which interpreted the substantiveness as widening the scope of eligible staff, found it difficult to decide whether staff at the very margin of the substantiveness test were PBRF-eligible.
- 280 Some TEOs reported that they did not feel able to exercise fully their discretion to decide PBRF-eligibility because they were not confident that their particular and untested interpretation of the test was accurate.
- 281 During 2003 some TEOs sought 'rulings' from the PBRF team on the application of the eligibility criteria to staff on the boundaries of the substantiveness test. TEOs reported that they received some advice from the TEC but found that to be inconsistent for individuals or groups of staff in like circumstances. This confirmed their view that the eligibility criteria were ambiguous and that a particular interpretation of the eligibility criteria might be possible.
- 282 Some TEOs in the study reported that they were not able to secure adequate clarification from the TEC about the application of the eligibility criteria to staff at the boundaries of the substantiveness test in the critical period around July and August 2003. This caused those TEOs to become very concerned about the interpretation of eligibility being applied by other TEOs.

The apparent re-emergence of the substantiveness test

- 283 Some participants in this evaluation reported that they had not recognised the importance of, or the difficulty in, applying the substantiveness test until it appeared in the eligibility flow diagram in the 25 July 2003, *Guidelines*. This appeared to be a consequence of:
- that fact that they had not noticing the test because of the volume of detailed information they received;
 - the 25 July *Guidelines* appearing to make the matter suddenly and obviously very important;
 - their understanding of the quality assessment process growing over the year such that by late June 2003 they needed more detailed explanations of key matters; and

- while they had applied the eligibility rules before the final *Guidelines* were issued (the wording had not altered from May) some TEOs believed that they were now facing more inclusive eligibility criteria than they had applied. This may reflect changes in TEO interpretation, understanding or tactical use of the substantiveness test.

Staff on leave, transferring, in concurrent employment and who change their employment status during the year

284 Some TEOs reported that they would appreciate an opportunity to review the eligibility criteria covering the following staff. TEOs are seeking to benchmark the interpretations they applied with an external moderator and to be reassured that all TEOs are applying the criteria consistently.

- Staff on leave.
- Transferring staff.
- Staff concurrently employed by two or more TEOs.
- Staff who change their employment status during the year.

Implementation of the Validation and Verification of Staff Eligibility Audit

285 The Ministry of Education undertook the Staff Eligibility Audit on behalf of the TEC. The audit compared the PBRF Census with the Ministry of Education's Single-Data Return (SDR). TEO staff calendars, web sites and telephone directories were used to provide additional evidence for assessing whether the *Guidelines* had been complied with.

286 The aim of the audit was to ensure 'that all eligible staff were actually included by participating TEOs in the 2003 Quality Evaluation and that no eligible staff were excluded' (Tertiary Education Commission, 2004a, p258).

287 The report on the design and conduct, results and conclusion of that audit are set out in Appendix C of the TEC Report.

288 *The 2003 Assessment* reported that:

The rules for staff eligibility are one of the cornerstones of the PBRF. A number of anomalies between the SDR and PBRF Census data were identified during the staff eligibility audit. However, satisfactory explanations were provided by TEOs when anomalies were brought to their attention. All concerns raised by panel

members about the eligibility of specific staff were investigated, and in most cases the concerns were found to be justified. A review of the rules for staff eligibility will be undertaken in preparation for the 2006 Quality Evaluations (Tertiary Education Commission, 2004a, p260).

- 289 On the basis of the quantitative and qualitative data available from the above returns and processes, the MOE and the TEC identified various apparent anomalies. The TEC reports that all TEOs subsequently asked provided 'reasonable explanations for the differences between the PBRF Census data and the Ministry's SDR data'(Tertiary Education Commission, 2004a, 259).

Diminished confidence in the Eligibility Audit

- 290 TEOs in the cases studied and other stakeholders consulted in the scoping phase of this evaluation, expressed concern about the interpretation and application of the eligibility criteria during the 2003 Quality Evaluation. Some TEOs reported having strong misgivings about the results of the 2003 Quality Evaluation for other TEOs.
- 291 Some TEOs reported that they expected that the Eligibility Audit to be conducted by the Ministry of Education would be searching and detailed. TEOs expected the Ministry to visit the TEO, to call for a sample of employment arrangements of staff deemed to be ineligible by the TEO, up to 10%, and to examine those very closely over a number of days.
- 292 Most participants reported that they experienced the conduct of the audit as light-handed and the Ministry as too readily satisfied with the assurances of TEOs that they had applied the substantiveness test as required. TEOs contrasted the eligibility audit with what was seen as the more systematic audit of research outputs.
- 293 By contrast, the TEC was concerned that the audit of NROs would be rejected by TEOs as overbearing and intrusive and doubted that TEOs would welcome two rigorous examinations in a short period. Moreover there could be no audit on the scale proposed for all participating TEOs, as there were no resources by way of people, time or funding, to carry it out.
- 294 The perception of TEOs in this evaluation was that the eligibility audit was not an adequate audit of the eligibility criteria or an adequate audit of the application of the substantiveness test in TEOs.
- 295 Even if there was no deliberate manipulation of the eligibility criteria or the substantiveness test in the 2003 Quality Evaluation, some TEOs believe that there was gaming by others and that some TEOs benefited from that gaming. This is why some TEOs continue to believe, despite



the TEC's assurances, that not all PBRF-eligible staff in all TEOs participated in the 2003 Quality Evaluation.

- 296 The TEC has indicated that it will review the eligibility criteria and we have suggested that some of the people in TEOs who applied the current criteria contribute to that review. Such people have 'street level' knowledge about the different practices that operate at the boundary between avoidance and evasion of the eligibility criteria.
- 297 Unless the TEC addresses the current uncertainties surrounding the eligibility criteria and commits to undertaking a more rigorous and effective eligibility audit in 2006, TEOs are likely to take steps to ensure that they are not disadvantaged by the 'sharp' practices of other TEOs.

Summary

298 Determining the eligibility of staff for inclusion in the 2006 Quality Evaluation remains fundamental for the integrity of the PBRF and for the confidence of participating TEOs.

299 We **recommend** that:

- the TEC review the Staff Participation Criteria via a working party that includes TEO people who will use the rules; and
- the TEC consider establishing a small (email based) eligibility reference group (1 TEC, 2 TEO ,1 MOE and 1 employment law specialist) available during Quality Evaluations to resolve eligibility questions by issuing agreed 'rulings'. If the proposed review of the current eligibility criteria resolves almost all issues prior to the 2006 round, this group may not be required but the function would be perceived by TEOs as helpful to the process.

300 TEOs and stakeholders in the evaluation were strongly of the view that the eligibility rules must be seen to be applied consistently by all participating TEOs. The PBRF Staff Eligibility Audit is both an assurance that the rules are being applied consistently and a transparent, public base for creating and maintaining confidence in the whole system.

301 We **recommend** that:

- the intended coverage of the **Staff Participation Criteria** (employment criteria) be clarified
- the **substantiveness test** be reviewed for its scope and application
- the **eligibility audit** be a more systematic and robust examination of a TEO's application of the eligibility criteria, and that



- 302 The MOE and the TEC have indicated that they intend to review the data gathering instruments (PBRF Census and Single Data Return) before the 2006 Quality Evaluation.
- 303 When the compulsory education sector moved to a staffing-based funding formula the MOE undertook detailed roll audits of some schools. A person with a very deep knowledge of the rules and the practices of school managers in stretching those rules, sat in the school until they had seen or accounted for every student. The audit was a demonstration to the whole sector that creative interpretations of the classification of students would be discovered.
- 304 Such a demonstration may assure the sector that the Staff Participation Criteria were being applied consistently and that any TEO might be examined in detail. The audit team could include one or more people from other TEOs in the audit team to assist with the establishment and dissemination of knowledge about the eligibility criteria in TEOs.
- 305 The TEC took a high trust approach to the application of the eligibility criteria in 2003. If it is possible to maintain that high trust in the sector then that is a much more desirable goal than developing intrusive or expensive audit mechanisms.
- 306 The merits of the MOE or the TEC undertaking a detailed audit of the application of the eligibility criteria in a small number of TEOs at each Quality Evaluation is a matter for the review of the eligibility audit.
- 307 There is general agreement from TEOs and MOE that the data gathering instruments (Census, SDR) were not adequate. The MOE has indicated that a review of the data to be gathered in the future and the best instruments to do so will be completed well before the 2006 Quality Evaluation.
- 308 TEOs wish to receive early advice about any changes proposed for data collection so that they can align their internal systems and strategies with the new requirements.
- 309 The eligibility criteria are likely to function more effectively if they are constructed by some of those people in TEOs who will use the criteria to make decisions about eligibility. TEOs were clear that they did not wish to launch a cycle of ever-smarter rules followed by sharper practices. Nor do they see value in trying to construct or enforce rules for all circumstances. Their preference is for the current good will between TEOs and the TEC to be reflected in a shared review and construction of new staff participation criteria.

Eligibility of TEOs

310 *Investing in Excellence* noted that:

All New Zealand-based degree-granting tertiary education providers, and all subsidiaries that are wholly owned by a New Zealand-based degree-granting tertiary provider, will be evaluated in the PBRF. This criterion reflects:

- the guiding principle of comprehensiveness
- the Government's desire to improve the quality of information on research output
- the need to align the PBRF with the Education Act requirement that degrees be 'taught mainly by people engaged in research (Ministry of Education and Transition Tertiary Education Commission, 2002, p11).

311 This is consistent with the Aims and Role of the Performance-Based Research Fund detailed in paragraphs 9 to 20 of *Investing in Excellence* (Ministry of Education and Transition Tertiary Education Commission, 2002, p7).

312 During our evaluation participants and stakeholders advised us that, in the development of a funding scheme to replace EFTS, it had been the intention and understanding of the tertiary sector generally, that all degree-granting tertiary education providers would participate voluntarily in the proposed PBRF.

313 As matters transpired, however, following the publication of *Investing in Excellence* a significant proportion of the institutes of technology and polytechnics decided that they would not participate in the proposed 2003 Quality Evaluation.

Election to participate

314 Participation in the 2003 Quality Evaluation was voluntary for TEOs. The proportion of participating TEOs might be seen as an indicator of the tertiary sectors' alignment with the overall PBRF policy intent. *The 2003 Assessment* reported that:

Of the 45 PBRF-eligible TEOs, 22 participated in the 2003 Quality Evaluation. The 22 comprised eight universities, two polytechnics, four colleges of education, one wānanga, and seven private training establishments (Tertiary Education Commission, 2004a, p1).

- 315 However, the 22 participating TEOs contained the overwhelming majority of all PBRF-eligible staff. Most of those staff were employed in a university. The non-participating TEOs were mainly small TEOs with few PBRF-eligible staff.
- 316 If the overall intent of the PBRF was 'to reward and encourage excellence...(in research) in all TEOs through the 2003 Quality Evaluation, then the responses of most polytechnics, some wānanga and some private training enterprises were not congruent with that objective as they elected not to participate. This does not mean that the policy will not succeed over time. It may be that TEOs that wish to secure funding from the PBRF in the future, will elect at some stage to participate. And it may be that TEOs, which did not participate in the 2003 Quality Evaluation, intend to prepare their staff to participate in a future Quality Evaluation.
- 317 Some participants in this evaluation regard the PBRF as very suited to the research and teaching activities of universities. One respondent commented, 'it was for universities really.'
- 318 Some TEOs in this evaluation considered that they would receive no benefit, financial or reputational, from participating in the 2003 Quality Evaluation. Their reasons for not participating included:
- polytechnics, colleges of education and institutions that focussed primarily on 'applied' research, were not confident that what they understood to be a fresh and broader definition of what was to count as research in the PBRF would be applied to the assessment of their EPs. They saw the peer review panels as dominated by traditional fields of research, by traditional research paradigms and by academics unlikely to be sympathetic or knowledgeable about the measures of excellence applying in their fields;
 - to achieve a Quality Category C, a researcher needed a research career during the six years prior to the 2003 Quality Evaluation. Many TEOs recognised that their staff could not meet this requirement for the 2003 Quality Evaluation. This was particularly the case for organisations that had only recently secured degree-granting status from NZQA. These TEOs reported that they saw no benefit in emphasising the gap between their staff and established academics to students, by categorising most of their staff, publicly, as 'Rs';
 - some TEOs calculated that the funding they might secure through the PBRF was unlikely to be greater than their costs of participation; and
 - an awareness that the Government was considering making additional funding available to the polytechnic sector and a

calculation that even if that was a small sum it was likely to be less costly to seek this funding than to participate in the PBRF.

319 Some TEOs without an established research culture or supporting infrastructure, but which are striving to build a research culture, entered the 2003 Quality Evaluation in the full expectation that they will never score so poorly again. They reported that they may reconsider their current strategy of seeking to compete with the established research cultures of universities if:

- they observe, on the basis of the results of the assessments, that their research activity is not measured appropriately. This is the view, considered in Chapter 5, that there were inherent advantages to traditional academic fields, forms and practices, in the 2003 Quality Evaluation;
- there is insufficient financial return; or
- the impact of poor results is judged to threaten the viability of the TEO.

320 Contributions to the OECD's programme on Institutional Management in Higher Education (IMHE) have noted that new and emerging TEOs, that are not traditional universities:

were not traditionally resourced for research and as a consequence have poor institutional infrastructure and technical support...Moreover, because academic staff at new institutions were hired originally to teach, they often lack the prerequisites – a postgraduate qualification, for example – and the necessary research experience....As the parent of many new disciplines, many of which had no research tradition, institutions face particular difficulties achieving recognition and funding, and navigating from successful applied and professional teaching programmes to research postgraduate activity(Hazelkorn, 2002, p6).

321 Stakeholders in the polytechnic sector reported that they saw the PBRF as an appropriate funding allocation mechanism for universities but not as an appropriate funding allocation mechanism for polytechnics and universities seeking the same funds. Polytechnics are seeking an alternative funding mechanism.

322 The PBRF was intended to integrate a funding allocation mechanism with a measure of research quality so as to stimulate excellence in research and lively research environments across the tertiary sector. The PBRF provided an opportunity for all TEOs to have the excellence of their research assessed, but all TEOs did not participate in the 2003 Quality Evaluation.



- 323 It is apparent that the substantive participants in the 2006 Quality Evaluation will be the universities. As universities currently contain most of the PBRF-eligible staff within the tertiary education sector this will mean that the policy intent will largely be met.
- 324 As indicated above, the government has agreed to provide funding to polytechnics to support some of their research activities with enterprises. This funding is understood, in the polytechnic sector, to be a small fund directed at the research needs of polytechnics and to be in part, an alternative to the PBRF.
- 325 The results of the 2003 Quality Evaluation have drawn attention to the relatively small number of research active staff in TEOs other than universities. TEOs may choose to enter future Quality Evaluations to assure themselves, stakeholders and the government that their research and teaching are closely interdependent, that the TEO meets international standards of research and teaching and that their degrees are being taught by people engaged in research, pursuant to the Education Act 1989.

Chapter 4: Direct and indirect costs of the implementation of the PBRF and the conduct of the 2003 Quality Evaluation

Introduction

326 The Performance-Based Research Fund (PBRF) Working Group emphasised that the proposed PBRF scheme should impose minimal administrative, compliance and assessment costs for providers, consistent with the creation, implementation and operation of a robust and credible system. That determination was recorded in an Efficiency principle.

Efficiency: administrative and compliance costs should be kept to the minimum consistent with a robust and credible process (Ministry of Education and Transition Tertiary Education Commission, 2002, p8).

327 The PBRF Working Group considered the costs of introducing the PBRF and noted that:

- introducing the PBRF would increase costs in the tertiary sector and the funding allocation system and there would remain a question, to be determined over time, as to whether the benefits justified the costs;
- the extent of the additional costs for participating TEOs would depend upon the degree to which providers were already taking steps to identify and raise their research quality. The understanding was that many providers were already implementing Research Assessment Exercises or Institutional Grant Scheme-style systems for internal research grant allocations;
- while the PBRF would increase administrative and compliance costs for providers, these higher costs were, in part, a result of developing a robust and credible system; and
- there was acceptance that the costs of the 2003 Quality Evaluation would include one-off costs for the MOE, the TEC and TEOs because a new scheme was being implemented for the first time.

328 The Tertiary Education Advisory Commission's Fourth Report (TEAC) noted that one of the strengths of the funding allocation mechanism that the PBRF was to replace was that:

the administrative and compliance costs associated with the allocation of Equivalent Full-time Students (EFTS) subsidies are low, and significantly lower than (per dollar allocated) than would



be the case under a performance-based funding system (Tertiary Education Advisory Commission, 2001c, p84).

- 329 The implementation of the PBRF as a funding allocation mechanism was a matter of policy. The rationale for the adoption of the PBRF scheme and the expected benefits, are set out in the reports of Tertiary Education Advisory Commission and the TEC leading to the implementation of the PBRF in 2003 (Tertiary Education Advisory Commission, 2000, 2001a, 2001b, 2001c). As the adoption of the PBRF was a matter of policy the benefits are assumed and are not considered in this Chapter.
- 330 The PBRF was understood to be a more expensive funding allocation mechanism than EFTS, but how much more expensive it may or may not prove to be is a matter for future retrospective evaluation, not prospective quantification in this evaluation.
- 331 Examining whether the PBRF will generate an allocation of extra research funding by the government and whether the PBRF process will generate an allocation of those funds between research providers that will generate research outcomes that are of higher quality than would have been generated by any other funding allocation mechanism (EFTS, contestable fund) is not the purpose of this Chapter.
- 332 For these reasons, the Efficiency principle is understood to be a requirement that the costs of the PBRF be minimised at least until it matures and a broader cost/benefit analysis can be carried out.
- 333 This chapter is in four sections that answer four broad questions:
- What were the administrative costs to MOE and the TEC and the compliance costs to TEOs from implementing the PBRF in 2003, operating it until 2006 and conducting the first Quality Evaluation?
 - What is likely to happen to the level of administrative costs to the MOE and the TEC and the level of compliance costs to TEOs when the PBRF is fully implemented after the 2006 Quality Evaluation and conducting the 2006 Quality Evaluation? And what does that mean for the PBRF system?
 - Is it possible to establish a benchmark for the level of administrative and compliance costs in the PBRF?
 - What did MOE, the TEC and TEOs say about the administrative and compliance costs in 2003 and about their expectations for 2006?

Limitations to the estimates of costs used in this Chapter

- 334 The MOE and the TEC estimates are drawn from Appropriations in the 2002 and 2003 Budgets (Table 1). Appropriations are monies government agencies are authorized to spend, not accounts of what they have spent.
- 335 The annual administrative costs of the PBRF to the MOE and the TEC are not known as the PBRF is only in its second year (2004). What funds might be appropriated in future Appropriations is also unknown.
- 336 The TEO estimates were supplied by nine TEOs. TEO estimates were not made on common assumptions or base lines or for the same purposes.
- 337 The nine TEOs employed 81% of PBRF-eligible staff (2003) and received 91% of the total PBRF funding allocated in 2004. Most TEOs who provided estimates, recorded their actual expenditure on PBRF project support, personnel, administration, IT, library and other services. All TEOs reported that staff time was a significant component of their costs but not all provided complete estimates of that cost.
- 338 The annual, ongoing costs to TEOs of their compliance with the PBRF are unknown.

'Administrative' and 'compliance' costs

- 339 The TEC manages the PBRF. The PBRF links the assessment of excellent research in the tertiary sector with a performance-based funding allocation mechanism.
- 340 There are three measures of research performance in the PBRF. A periodic Quality Evaluation using expert panels to assess research quality based on material contained in Evidence Portfolios (EPs) submitted by individual researchers; a measure for research degree completions (RDC) by a TEO; and a measure for the external research income (ERI) secured by a TEO. RDC and ERI are annual measures and apply from 2003.
- 341 The PBRF was implemented in 2003 and will complete a transition phase to full implementation by 2006. The PBRF is now in place and will impose annual administrative costs on MOE, the TEC and compliance costs on TEOs; the level of those costs is not known.
- 342 The initial Quality Evaluation was prepared and conducted in 2003. Further Quality Evaluations will be conducted in 2006 and 2012. Each Quality Evaluation will impose periodic administrative costs on the MOE and the TEC and periodic compliance costs on TEOs; the level of those costs is not known.

343 In this chapter, annual costs incurred by the MOE and the TEC in administering, monitoring and reporting on the PBRF are termed 'administrative costs'. Annual costs incurred by TEOs in gathering and submitting data returns to the TEC or MOE for the PBRF (RDC and ERI) and the periodic costs of all of the tasks associated with preparing for and participating in a Quality Evaluation, are termed 'compliance costs'.

Estimating the costs of implementing the PBRF, conducting the 2003 Quality Evaluation and administering the PBRF for the period 2004-2006

Administrative costs to the MOE and the TEC – 2002 to 2006

344 The operating expenditure appropriated by Parliament for the MOE and the TEC is set out in Table 1. These are best first estimates only of the costs of implementing and administering the PBRF and the 2003 Quality Evaluation in the period 2002-2007 for the MOE and the TEC.

Table 1: Estimated Operating Expenditure for the implementation, operation and review of the PBRF for the MOE and the TEC for the period 2002 to 2007.

		2002/03 (\$million)	2003/04 (\$million)	2004/05 (\$million)	2005/06 (\$million)	2006/07 (\$million)
TEC	Budget 02	0.05	0.192	0.204	0.174	0.174
	Budget 03	1.916	2.881	0.845	0.845	0.845
	Total	1.966	3.073	1.049	1.019	1.019
MOE	Budget 02	0.474	0.122	0.028	0.028	0.028
	Budget 03	-	0.270	0.430	0.350	0.270
	Total	0.474	0.392	0.458	0.378	0.298
Total		2.44	3.465	1.507	1.397	1.317

Source: Budget Appropriations 2003, Ministry of Education

345 The data in Table 1 indicates that \$5.905 million was budgeted for the 2002/03 and 2003/04 years for MOE and the TEC to implement the PBRF and conduct of the first Quality Evaluation.

346 Table 1 indicates that costs of approximately \$1million are budgeted annually for the TEC in the years 2004/05, 2005/06 and 2006/07 to:

- complete the 2003 Quality Evaluation;
- review the implementation of the PBRF;



- identify and undertake any modifications to the PBRF;
 - administer the PBRF in the years 2004 to 2006; and
 - prepare and conduct the 2006 Quality Evaluation.
- 347 The costs of conducting the 2006 Quality Evaluation will be considerably greater than the estimates in Table 1 provide. The costs of the peer review panel process for the 2003 Quality Evaluation are estimated at approximately \$1.1 million. If any proposals to alter the composition and conduct of the peer review panels and their assessment processes are adopted, this will increase the administrative costs of the 2006 Quality Evaluation to the TEC. Whether the overhead costs of the implementation of the PBRF were allocated properly is not considered in this analysis.
- 348 The data in Table 1 provides for decreasing provision for the administrative costs of the MOE from \$458,000 for 2004/05, \$378,000 for 2005/06 and \$278,000 for 2006/07.
- 349 Other Crown agencies (e.g. Treasury, Crown Law) will have incurred costs arising from the implementation of the PBRF that were unforeseen or not funded by any Appropriation; these costs have not been identified or quantified.

Compliance costs to TEOs – PBRF implementation and the 2003 Quality Evaluation

- 350 The data in Table 2 is based on estimates provided by nine TEOs. The estimates provided included records of direct expenditures, and for six TEOs, an estimate of the cost of staff time. For two of the remaining three TEOs WEB Research estimated the cost of staff time on the basis of one day per EP per PBRF-eligible staff member¹; at the rate of \$80,000 per FTE. This rate was slightly less than the FTE rate used by the two largest TEOs that provided estimates. In the third case, a small TEO believed it could not distinguish staff costs from its ongoing business and WEB Research provided an estimate.
- 351 TEOs that participated in the evaluation or who provided estimates of costs were assured of their anonymity. It is impossible to provide a full account of all figures and assumptions and preserve that anonymity in such a small tertiary sector. The publication of *The 2003 Assessment* with its detailed tables of funding and participation made preserving TEO anonymity in this Chapter even more difficult.
- 352 To reflect the substantial uncertainty about the estimates of the administrative and compliance costs, this analysis constructs a:

¹ PBRF-eligible staff are employees of TEOs who meet the Staff Participation Criteria of the Guidelines and are required to submit an EP.



- Low;
- mid-range; and
- high.

estimate of costs for the 2003 and 2007 PBRF rounds.

Table 2: Estimated costs of implementing the PBRF and participating in the 2003 Quality Evaluation to TEOs for the period 2003-2004.

		(\$million)
Total eligible staff in TEOs that provided data	6459	
Percentage of all PBRF-eligible staff in TEOs that provided data	81%	
Costs estimated by TEOs extrapolated to 100% PBRF-eligible staff		8.200

353 The total actual cost estimated by the nine TEOs, including staff time, was \$6.650 million. This estimate represented 81% of 2003 PBRF-eligible staff (6459) who received 91% of the PBRF funding allocated in 2003 and was extrapolated to provide an estimate for 100% of PBRF-eligible staff, a total of \$8.200 million. This provides the **low** estimate of total costs for all TEOs in 2003 used in Table 3.

354 In the UK the manger of the RAE extrapolated the costs of one Higher Education Institution (HEI) to provide an estimate of the costs of the RAE for all HEIs. Using the same method the detailed estimate of costs provided by one medium sized TEO was extrapolated to provide an estimated total cost for all TEOs in 2003 of \$12.200 million; this provides the **mid-range** estimate in Table 3.

355 Extrapolating the highest estimate of costs provided by a university and the highest estimate of costs provided by a TEO that is not a university provided an estimate of costs for all TEOs in 2003 of \$15.400 million; this provides the **high** estimate in Table 3.

Low, mid-range and high estimates of total administrative and compliance costs – MOE, TEC, TEOs – for 2002/2003 only

356 The data in Table 3 combines the costs of the implementation of the PBRF and the conduct of the 2003 Quality Evaluation to the MOE, the TEC (Table 2) and TEOs (Table 3) to provide low, mid-range and high estimates of the total administrative and compliance costs in 2003.

357 The total administrative and compliance costs (MOE, TEC, TEOs) of the implementation of the PBRF in 2002/03 and the conduct of the 2003 Quality Evaluation are expressed as a percentage of the proportion of



the PBRF funding for the 2004-2006 period that was allocated to TEOs on the basis of the 2003 Quality Evaluation. This provides a basis for future comparison and a measure, for TEOs, of the efficiencies and incentives for them within the PBRF system.

358 The PBRF Indicative Funding for the period 2004 to 2006 is:

- 2004 - \$18.2 million;
- 2005 - \$43.613 million; and
- 2006 - \$105.826 million,

a total of \$167.639 million. **60%** of that funding, \$100.583 million, was allocated for the 2004-2006 period on the basis of TEO performance in the 2003 Quality Evaluation. (*Source Table 8.3, The 2003 Assessment*)

Table 3: Estimated costs of the implementation of the PBRF and the conduct of the 2003 Quality Evaluation to the MOE, the TEC and TEOs.

	(\$million)	TOTAL (\$million)	Percentage of QE Funding 2004-2006
Total TEC & MOE costs 2002 - 2004	5.905		
Total TEO costs 2003			
Low	8.200	14.105	14%
Mid-range	12.200	18.105	18%
High	15.400	21.305	21%

Notes:

- i. *Total* is estimate (low, mid-range, high) plus TEC & MOE costs
- ii. *Percentage* is *Total* as percentage of \$100.583 million.

359 The data in Table 3 indicates that the total administrative and compliance costs of implementing the PBRF in 2003 and conducting the 2003 Quality Evaluation is likely to have ranged between \$14 and \$21 million dollars or between 14% and 21% of the total PBRF funding for the period 2004-2006 allocated by the 2003 Quality Evaluation.

360 This is, on any measure, a high cost. However, the combination of chosen design elements, complexity, newness and demands on staff time ensured that the costs of the implementation of the PBRF and the simultaneous conduct of the 2003 Quality Evaluation were always going to be high. They were also high as a proportion of the PBRF funding allocated because the PBRF funding allocated by the 2003 Quality Evaluation and the RDC and ERI results, was the smallest it may ever be.



361 No funding was appropriated or provided by the Crown to TEOs to compensate them for costs they incurred in complying with the PBRF and the 2003 Quality Evaluation.

Administrative costs to the MOE and the TEC from administering the PBRF in the period 2004-2006

362 The one-off and sunk costs to the MOE and the TEC associated with the preparation and implementation of the PBRF in 2003 are reflected in their respective budgets for the years 2002-2004 (Table 1).

363 A complete summary of the costs of the implementation of the PBRF must include the administrative costs to MOE and the TEC and compliance costs to TEOs of the PBRF for each of the years 2004-2006.

364 In the years between Quality Evaluations there is a range of ongoing costs to the TEC associated with the administration of the PBRF. Those costs will include:

- preparation and issuing of returns to TEOs;
- audit and checking with TEOs on data on RDC and ERI submitted to MOE;
- analysis of those returns;
- funding adjustments and transactions;
- monitoring and evaluating the ongoing impact of the PBRF;
- policy analysis and development;
- consultation and liaison with the sector, especially in the period before the next Quality Evaluation; and
- preparing an annual report on the PBRF.

365 There will be additional administrative costs to the MOE and the TEC in the period 2004 to 2006 as the PBRF is reviewed, altered and prepared for the planned 2006 Quality Evaluation. Cabinet intends that the second Quality Evaluation take place in 2006 in order to benefit from the experiences of the 2003 Quality Evaluation.

366 The MOE and the TEC have indicated that they intend to review the adequacy of the existing data sets they hold about TEOs and the efficiency of the collection instruments (RDC, ERI, PBRF Census) associated with the administration of the PBRF. Any changes in those are expected to result in very small increases in the administrative and compliance costs to TEOs.



367 Table 1 indicates that the estimated administrative costs to the MOE and the TEC for the PBRF in the years 2004 to 2006 are \$2.744 million. The costs of the 2003 Quality Evaluation indicate that the Table 1 estimates of the costs to the MOE and the TEC to prepare and conduct the 2006 Quality Evaluation are too low by at least \$1 million; the approximate cost of the 2003 peer review panels. The estimate of the costs to the MOE and the TEC of administering the PBRF in the period 2004-2006, the transition period and the preparation of the 2006 Quality Evaluation has been increased to \$3.744 million for this analysis.

Costs of complying with the PBRF to TEOs in the period 2004-2006

- 368 A complete summary of the costs of the implementation of the PBRF must also include the compliance costs to TEOs associated with the administration of the PBRF for the period 2004-2006.
- 369 The cost of gathering and providing information on RDC and ERI is the major compliance costs to TEOs in the years between Quality Evaluations. These costs are expected to remain proportionately small even if the MOE and the TEC require additional information or institute new data collection methods or collection instruments. WEB Research has estimated the costs for all TEOs in the PBRF as \$1 million a year for each of the years 2004 to 2006, a total of \$3 million.
- 370 The total costs of administering and complying with the PBRF in the period 2004-2006 to MOE, the TEC and TEOs is estimated at \$6.740 million.

Estimating the total costs of the first complete cycle of the PBRF in the period 2002-2006

- 371 Table 4 adds the costs of implementing the PBRF, conducting the 2003 Quality Evaluation (Table 3) and the costs of administering and complying with the PBRF in the period 2004-2006 (\$6.740 million) to provide an estimate of the total costs of the implementation and first complete cycle of the PBRF.
- 372 The costs are expressed as a percentage of the total PBRF funding to be allocated to participating TEOs in the period 2004 to 2006; \$167.639 million. This provides a basis for future comparison and a measure of the efficiency by which the PBRF system allocates the total funding in the period against the total costs for that PBRF round.

Table 4: Estimated total costs of implementing the PBRF, conducting the 2003 Quality Evaluation and administering and complying with the PBRF in the transition period 2004-2006 to the MOE, the TEC and TEOs.

	Cost MOE, TEC TEOs 2002 -04 (\$million)	Costs MOE,TEC TEOs 2004 - 06 (\$million)	TOTAL MOE,TEC TEC 2002-2006 (\$million)	Percent of Total PBRF Funding 2004-2006
Low	14.105	6.744	20.849	12%
Mid-range	18.105	6.744	24.849	15%
High	21.305	6.744	28.049	17%

373 The data in Table 4 indicates that the estimated total costs of implementing the first complete cycle of the PBRF ranged between \$21 and \$28 million dollars. This was between 12% and 17% of the total PBRF funding allocated for the period 2004-2006.

374 TEAC noted that the New Zealand Vice-Chancellors' Committee had sought to estimate the likely central administrative costs of the proposed PBRF:

According to the New Zealand Vice-Chancellors' Committee, the central administrative costs (including auditing costs) associated with the mixed model are likely to be under \$0.5m per annum (with higher costs once every five years as a result of the quality rating exercise). There will, of course, be additional costs for those providers that seek funding through the PBRF (the costs of academic staff assessment and moderation, the collection and provision of additional information, the preparation of Research and Research Training Management Plans etc)(Tertiary Education Advisory Commission, 2001c, p101).

375 The data in Table 4 indicates that the costs will not approach that level before 2007.

376 In a later part this report considers what might be an appropriate level for the administrative and compliance costs of the PBRF by considering the administrative and compliance costs of the UK and Hong Kong RAEs. The administrative and compliance costs for the last UK RAE are estimated at 1% of the total funding allocated. The comparable figure for the Hong Kong RAE was less than 1%.

377 The data in Table 4 indicates the total costs in New Zealand were a great deal higher percentage of the total performance-based funding allocated for the period 2004-2006 than were the administration and compliance costs for the UK and Hong Kong RAEs.



Estimating the costs of conducting the 2006 Quality Evaluation and administering the PBRF in the period 2007-2012

- 378 A number of factors will influence the level of administrative and compliance costs of the 2006 Quality Evaluation. The review of the 2003 Quality Evaluation may alter aspects of the assessment, reporting or funding processes. There may be more peer review panels or different sorts of panels, if more staff participate in 2006 there may be higher costs for panellists time. There may also be savings from the all participants understanding the requirements better and the TEC and TEOs having improved systems in place to prepare for and conduct the 2006 Quality Evaluation.
- 379 The estimate of the costs of the 2006 Quality Evaluation in Table 5 are based on the costs estimated for the MOE, the TEC and TEOs for the 2003 Quality Evaluation only, and as set out in Table 3.
- 380 The estimate for the costs of the 2003 Quality Evaluation included an estimate of the costs of the implementation of the PBRF to MOE and the TEC of approximately \$2.500 million. This is a cost that ought not to recur in 2006 because the PBRF has now been implemented.
- 381 However the estimate for the costs of the 2006 Quality Evaluation in Table 5 includes a similar provision for \$2.500 million. This is because the estimated Appropriations (Table 1) for the costs of the 2006 Quality Evaluation to MOE and the TEC do not reflect the likely costs of conducting the 2003 Quality Evaluation.
- 382 It is also the case that the MOE and the TEC will be better prepared for the 2006 Quality Evaluation, and that a proportion of the costs of preparing for the 2006 Quality Evaluation, incurred largely in one year in 2002/03, will be spread over the years 2004-2006. For this reason the estimated costs of conducting the 2006 Quality Evaluation to the MOE and the TEC have been reduced by 20% from \$5.905 million to \$4.724 million dollars.
- 383 The costs estimates provided by TEOs showed staff time as ranging between 65-70% of their total estimated costs. All TEOs in the evaluation reported that they expected to make substantial reductions in the costs of staff time in a Quality Evaluation by the time of the 2006 Quality Evaluation. If this is the case then the compliance costs to TEOs will reduce very substantially.
- 384 This evaluation accepts that TEOs will be much better prepared for the 2006 Quality Evaluation and that there will be more effective management of the research data required to complete EPs. This evaluation reduces the estimated compliance costs to TEOs of the 2003 Quality Evaluation (Table 3) by 20% to provide a range of estimates of the compliance costs of the 2006 Quality Evaluation to TEOs.



385 As was the case for the estimated costs of the 2003 Quality Evaluation, the costs of the 2006 Quality Evaluation to TEOs are expressed as a percentage of the PBRF funding allocated to TEOs by the 2006 Quality Evaluation.

386 In the period 2007 to 2012 the funds projected to be allocated by the PBRF will increase to \$185 million (2007) a year. The ratios that emerge in this analysis are entirely dependent upon that funding being allocated. Increases in student enrolments, higher research degree completion, greater external research income secured and possible increases in government funding for excellence in research may increase the funds available.

387 The report estimates that there will be approximately \$200 million a year funded for the six years to 2012, a total of \$1.2 billion dollars for the period 2007-2012. The estimate is that TEOs will receive **60%** of the \$1.2 billion allocated for the period 2007-2012, a sum of \$720 million.

Table 5: Estimated costs of administering and complying with the 2006 Quality Evaluation to the MOE, the TEC and TEOs.

	Costs 2003 QE -Table 3 (\$million)	Costs 2003 QE less 20% (\$million)	TOTAL (\$million)	60% of total PBRF Funding 2007-2012
MOE and TEC	5.905	4.724		
Low	8.200	6.560	11.284	1.57%
Mid-range	12.200	9.760	14.484	2.01%
High	15.400	12.320	17.044	2.37%

388 The percentages for the costs of the 2006 Quality Evaluation are lower than the percentages for the 2003 Quality Evaluation in large part because of the growth of the PBRF funding allocated. Whether there is a real reduction in administrative and compliance costs for the 2006 Quality Evaluation is a matter for the TEC to monitor.

389 The data in Table 5 suggest that if the TEOs' costs of participation in the 2006 Quality Evaluation are reduced by 20% through efficiencies at the center and at the level of TEOs, then the proportion of costs to funding allocated on the basis of the 2006 Quality Evaluation (60%) will fall from 14 to 21% (2003 Quality Evaluation - Table 3) to between 1% and 2.3% for the period 2007 to 2012.

390 These are certainly lower percentages and, within the strong limitations of this analysis, they suggest that the costs of the 2006 Quality Evaluation may be much lower as a proportion of the PBRF funds



allocated in the period 2007 – 2012 than were the proportion of costs of the 2003 Quality Evaluation.

Administrative costs to MOE and the TEC from administering the PBRF in the period 2007-2012

391 The report estimates the ongoing costs of the administration of the PBRF for the years 2007 to 2012 will be \$1 million per year for the TEC. The costs to the MOE are likely to be smaller by 2012 as the policy is resolved; we have estimated the annual ongoing cost to MOE as \$0.250 million per year. The estimate of the administrative costs to the MOE and the TEC for the period 2007 to 2012 is \$7.500 million dollars.

Costs of complying with the PBRF to TEOs in the period 2007-2012

392 On the basis of the figures supplied by TEOs the report estimates total compliance costs, for all participating TEOs, as \$1 million per year for the period 2007-2012; a total of \$6 million dollars.

393 This allows approximately \$45,000 dollars for each of the 22 TEOs that participated in the 2003 Quality Evaluation, for each year in the period 2007-2012. No attempt is made to refine this estimate further.

Estimating the total costs of the second complete cycle of the PBRF in the period 2007-2012

394 This analysis assumes that by 2007 there will be some reduction in the ongoing administrative and compliance costs of the PBRF to the MOE, the TEC and TEOs. This is likely to come from better planning, less haste, improved data gathering instruments (PBRF Census, SDR and any new instrument), improvements in ongoing record keeping in TEOs and better understandings between parties about how the whole PBRF might be made more efficient.

395 On the basis that there will be some reduction in the ongoing administration and compliance costs of the PBRF compared with the costs in 2003, the estimated costs in Table 5 have been reduced by 20% for Table 6.

396 The total estimate for MOE and the TEC in the period 2007-2012 is \$6.000 million and for all TEOs over the same period, \$4.800 million. The total estimate of the ongoing administrative and compliance costs of the PBRF for the period 2007-2012 to the MOE, the TEC and TEOs, is therefore \$10.800 million

397 There remain many factors that could alter the actual costs so these estimates are of limited reliability as indicators of reducing costs. It is

the case that the increased funding available from 2007 will reduce the percentage that administrative and compliance costs are as a proportion of the total funds allocated even if those costs remain the same.

398 The data in Table 6 combines the estimated costs of the 2006 Quality Evaluation to the MOE, the TEC and TEOs with the estimated ongoing costs of the PBRF for the period 2007 to 2102 for the MOE, the TEC and TEOs and expresses that cost as a percentage of the total PBRF funding allocated in the period 2007 to 2012 (\$1.2 billion).

Table 6: Estimated total costs of conducting the 2006 Quality Evaluation and administering and complying with the PBRF in the period 2007-2012 to the MOE, the TEC and TEOs

	Costs of PBRF for the period 2007-2012 (\$million)	TEO costs of 2006 QE	TOTAL (\$million)	Costs as percentage of total PBRF funding allocated 2007-2012
MOE, TEC & TEOs costs of PBRF for 2007-2012	10.800			
TEO costs of 2006 QE				
Low		6.560	17.360	1.45%
Mid-range		9.760	20.560	1.71%
High		12.320	23.120	1.93%

399 The data in Table 6 indicates that the percentage of administrative and compliance costs as a proportion of the total funds allocated for the period 2007-2012 is likely to fall from a range between 12% to 17% (Table 4) to a range closer to between 1.5% to 2% for the period 2007-2012.

Discussion

400 This Chapter is not a cost-benefit analysis. All it may do is emphasise where the costings identify sensitivities that will need to be managed as the PBRF and TEO behaviours mature over the next years. It may help to refine the development of the performance-based research funding method chosen over other funding delivery methods.

401 The analysis suggests that:

- provided costs remain under control, they will fall to a low percentage of the PBRF funds distributed; though that is
- provided that the amounts of funds being distributed through the PBRF increase at least as anticipated.

- 402 The analysis assumes that the government will remain committed to the PBRF and use it to increase research funding. Whether that may or may not occur is outside the scope of this review. It is also clear that the burden of costs in the PBRF is very heavily determined by the funding to be introduced by the government in the future and that this is not controllable. Not even the design element that expects excellence to be rewarded will necessarily leverage additional funding out of an unwilling government.
- 403 There is a risk that the incentives to control costs in the PBRF may not be symmetric. Within the PBRF system there is no advantage in 'gaming' the allocation of funds as that does not increase government funding to the system as a whole. The incentives for the system are to standardise and simplify processes so as to minimise administrative and compliance costs; this is consistent with the Efficiency principle.
- 404 However there appear to be incentives emerging for TEOs to game the allocation system. Table 7 indicates that there are strong returns to participating TEOs for their compliance and participation expenditures.

Table 7: Estimated average dollar return to nine TEOs for each estimated dollar spent on their participation in the PBRF and compliance with the 2003 and 2006 Quality Evaluations.

	TEO1 (\$)	TEO2 (\$)	TEO3 (\$)	TEO4 (\$)	TEO5 (\$)	TEO6 (\$)	TEO7 (\$)	TEO8 (\$)	TEO9 (\$)
2004-2006	13.00	2.00	0.80	4.50	14.00	12.00	13.00	12.00	19.00
2007-2012	79.00	13.00	5.00	28.00	90.00	74.00	78.00	74.00	117.00

Notes:

- i. Costs and PBRF funding for period 2004/06 are taken from Table 4.
- ii. Costs and PBRF funding for period 2007/12 are taken from Table 6.
- iii. PBRF funding for 2007-2012 was allocated to the nine TEOs in the same proportions as the PBRF funding each secured in the 2003 Quality Evaluation.
- iv. Because of the variations in the methods and assumptions applied by TEOs in estimating their costs the figures in Table 7 are very tentative indeed. Further, TEOs that participated in the evaluation or who provided estimates of costs were assured of anonymity. It is impossible to provide a full account of all figures and assumptions and preserve that anonymity for such a small tertiary sector. The publication of *The 2003 Assessment* with its detailed funding and participation tables made preserving TEO anonymity even more difficult.



- 405 The nine TEOs that provided cost estimates used in Table 7 represented 81% of PBRF eligible staff and secured 91% of the PBRF funding allocated in 2004.
- 406 The data in Table 7 indicates that for those nine TEOs, the average return on each dollar expended on the PBRF in the period 2003 to 2006, ranged between 0.80 cents and \$19.00 dollars.
- 407 The data in Table 7 indicates that for those nine TEOs the average return on each dollar expended on the PBRF in the period 2007 to 2012, including the 2006 Quality Evaluation is estimated to range between \$5.00 dollars and \$117.00 dollars.
- 408 The two lowest returns were to small TEOs with new or developing research cultures. The largest return was to a large TEO with an established research culture and a research support system. The second largest return was to a medium sized TEO that provided the lowest estimated costs for medium and large TEOs. The estimate of costs for this TEO was eight times less than the estimate for a similar sized TEO. This meant that given its low costs, the PBRF funding it received resulted in a higher average return than a comparable TEO.
- 409 A further indicator of possible incentives to TEOs to game is the reducing percentage compliance costs will be of the total funding allocated to the nine TEOs for the 2004/2006 period and estimated for the 2007/2012 period.
- 410 The data in Table 8 indicates that while the larger TEOs had differing levels of costs and differing levels of funding, by 2007, if their costs and the proportion of PBRF funding allocated to them remain much the same, then large TEOs are likely to see their costs as a proportion of the total funds allocated approach 1%.
- 411 Smaller TEOs had lower costs than the larger TEOs in 2003 but they also received very small amounts of funding. Again, if their costs remain much the same and their funding increases remain proportionate in 2007, then their costs as a proportion of the total funds allocated will decrease but not to the level of the larger TEOs.
- 412 What level of costs as a proportion of the total funds allocated is tolerable for a TEO may emerge in the period up to or following the 2006 Quality Evaluation. This is a key issue for TEOs that did not score well or that may be considering whether to continue or enter the PBRF.

Table 8: Estimated costs as percentage of total PBRF funding received by nine TEOs for the period 2004-2006 and the period 2007-2012.

	TEO1	TEO2	TEO3	TEO4	TEO5	TEO6	TEO7	TEO8	TEO9
2004-2006	8%	49%	130%	22%	7%	8%	8%	8%	5%
2007-2012	1%	8%	21%	4%	1%	1%	1%	1%	0.8%

Notes:

- i. Costs and PBRF funding for period 2004/06 are taken from Table 4.
- ii. Costs and PBRF funding for period 2007/12 are taken from Table 6.

413 This analysis has assumed that there are no advantages to TEOs from developing a range of strategies or behaviours, for example special preparation, presentation or employment practices, for the ongoing PBRF or Quality Evaluations. If this is valid and the compliance costs to TEOs can be bound, then the cost estimates are valid and the 20% savings may be possible.

414 It is too soon to know what factors will cause TEOs to decide that gaming the allocation of funds will pay and whether it is an action they are prepared to take, knowing how other TEOs will inevitably respond.

415 This analysis indicates that those who design, monitor and administer the PBRF, and the participants, will need to manage the system so that:

- o The incentives for the PBRF system as a whole are to standardise and simplify processes to minimise costs; and
- o The incentives to TEOs to spend to the limit of the expected marginal return to compliance costs are also minimised.

Establishing a possible benchmark for the level of administrative and compliance costs

416 The estimate of costs to the MOE, the TEC, TEOs indicates that the costs of the implementation of the PBRF and especially the 2003 Quality Evaluation were high as a proportion of:

- o the funding allocated on the basis of the 2003 Quality Evaluation; and
- o the total funding that will be allocated by the PBRF in the transition period, 2004-2006 (approximately \$170 million).

- 417 The PBRF replaces an existing mechanism for delivering funding to TEOs. The PBRF Working Group recognised that a performance-based funding delivery mechanism would more expensive (less efficient) than the EFTS funding delivery mechanism. The effectiveness of the PBRF will turn in part upon the degree to which it promotes a lively and productive research culture and that it does so at minimum cost.
- 418 But what might be an appropriate level of administrative and compliance costs for the PBRF could not be known until the PBRF was fully implemented. Some guidance might be by establishing benchmarks with other performance-based funds and a model that allocates funding on the basis of competitive bidding.
- 419 Limited comparisons can be made with the proportion of administrative and compliance reported by other performance-based funds such as the UK and Hong Kong RAEs. Those comparisons help to identify a possible benchmark for the administrative and compliance costs of the PBRF in 2007.
- 420 There are complexities in estimating the administrative and compliance costs of the UK and Hong Kong RAEs. In the ‘...first review of research policy and funding to have been conducted in England for many years...’ which reviewed the 1996 RAE, the Higher Education Funding Council for England (HEFCE) reported that:

The RAE is sometimes criticised as an expensive exercise. Relatively speaking, it is not. Returns from all institutions after the last exercise [1996] suggest that it cost slightly under £30 million. A more in-depth analysis of costs in one institution extrapolated to the sector as a whole suggests an upper limit for the cost of the RAE of £37.5 million, including opportunity costs. This represents just 0.8 per cent of the total funds allocated on the basis of the results of the exercise (Higher Education Funding Council for England, 2000, p19).

- 421 In a report on the 2001 RAE, the RAE Manager noted that the initial operating budget for the 2001 RAE was set in 1997 at £3.602 million. The actual cost was £5.100 million, an increase of 68% over the previous RAE in 1996. The RAE Manager observed that:

It is noticeable that many of the cost increases relate to improvements in the RAE process since 1996, including the participation of users, the establishment of large sub-panels in medicine, mandatory cross-referral for interdisciplinary research and the introduction of Non-UK-Based Advisers (Rogers, 2002, p19).

- 422 The £5.100 million is the estimated administrative cost, it does not include compliance costs to Higher Education Institutions (HEIs). The 2001 RAE Manager estimated that the compliance costs for HEIs were likely to be 50% higher than in 1996 (£37.5 million), an estimate for 2001 of £67 million. The increase was due to inflation and, amongst other things, increased requirements on HEIs to submit additional information and undertake more rigorous and demanding assessment processes prior to providing submissions.
- 423 Allowing for additional compliance costs to all HEIs over the extra two years of £5 million, brings the total estimated administrative and compliance costs of the 2001 RAE to £77 million.
- 424 The 2001 UK RAE was intended to allocate funding for five years. The 2006 RAE has been delayed until 2008 so the total funding will now be allocated over at least seven years. In the meantime the size of the fund has been increased significantly to allocate approximately £10 billion over the seven year period.
- 425 These estimates, given the many qualifications, indicate that the administrative and compliance costs (£77 million) of the 2001 RAE were approximately 0.8% of the total funds allocated (£10 billion). Administrative costs (£5.1 million) were approximately 0.05% of the total funding allocated.
- 426 The Hong Kong Universities Grants Committee (UGC) undertakes the Hong Kong RAEs. Boston (Boston, 2002, p84) reports that:
- four employees of the UGC undertook the bulk of the administrative work for the 1999 RAE. We have estimated that cost as HK\$1million;
 - this was regarded as barely adequate;
 - only overseas experts were paid; and that
 - the administrative costs of the panels was HK \$3 million.
- 427 There has been no estimate of the compliance costs imposed on the eight universities in the scheme. Boston reports that:
- ..in response to the RAEs, each university has tended to increase the number of staff with responsibilities for research administration. Further some of the universities have instituted their own internal RAEs (with inevitable administration costs). Additionally, of course, there are significant costs associated with the peer review system, including the preparation of submissions, the work of the panel chairs and members (the majority of whom are academics at Hong Kong universities), the preparation of



Research Strategy Statements, and so forth.(Boston, 2002, p84).

- 428 However the compliance costs will be lower on aggregate per FTE staff member than the PBRF because, among other differences, participants had only to select and supply up to five research outputs, there was no need to complete EPs nor were there internal panels. We have estimated the compliance cost for each university as being HK \$500,000 dollars, a total for the eight universities of HK\$4million.
- 429 The total estimated administrative and compliance costs for the 1999 Hong Kong RAE is \$10 million. We have included an estimated HK\$2 million in compliance costs incurred by all universities for the period to 2006.
- 430 The 1999 RAE in Hong Kong was intended to allocate funding for the triennium 2001-2004, however that has been extended as the proposed 2005/06 RAE has been delayed. In 2002 the RAE component of university funding was approximately HK\$3 billion suggesting that the 1999 RAE is likely to be used to allocate at least HK \$15 billion (over five years). It is likely that the funding will increase in the period up to the next RAE so the estimate of HK\$15 billion is conservative.
- 431 In New Zealand, the Foundation for Research, Science and Technology (FRST) allocates funding using a competitive bidding model. That model has administrative and compliance costs. To compete for FRST funds, organisations must prepare and submit detailed and substantial proposals that are evaluated by panels selected and supported by FRST.
- 432 FRST was not able to provide estimates of the compliance costs to bidding organisations. Universities and Crown Research Institutes, which are the largest group of bidders, employ staff full-time to prepare and manage their bidding processes. For the purpose of this analysis it is assumed that the level of costs to FRST are those of a mature scheme wherein costs have become lower over time.
- 433 FRST advised that it receives 3.5% to 4.5% of the funds it allocates as the administrative cost of the management and conduct of the bidding process. Across all funds and the organisation, FRST estimates that the average administrative costs of a competitive fund are approximately 3% of the total funds allocated.
- 434 If the compliance costs to organisations were 1% of the funds allocated, then the total administrative and compliance cost of allocating funds using a competitive bidding model is 4% of the total funds allocated.
- 435 From 2007 onwards, the PBRF will be fully implemented, with a total annual allocation of approximately \$185 million (on current estimates). Assuming that the government provides some new funding, the PBRF is likely to allocate around \$1.2 billion in the six years following the 2006



Quality Evaluation. On this basis, the administrative and compliance costs of the PBRF will be much lower as a proportion of the total funds allocated than will be the case during the three year transition period, 2004-2006.

- 436 These estimates, given the many qualifications, indicate that the administrative and compliance costs (HK\$10 million) of the 1999 RAE were approximately 0.07% of the total funds allocated (HK\$15 billion). Administrative costs (HK\$4 million) were approximately 0.03% of the total funding allocated.
- 437 In the course of this evaluation some TEOs suggested that their costs would be lower in the 2006 Quality Evaluation. The MOE and the TEC have indicated that they will review the implementation and operation of the PBRF in time for the 2006 Quality Evaluation; the review and some of the recommendations in this report may reduce the administrative and compliance costs of the PBRF.
- 438 Estimates of costs supplied by the MOE and some TEOs indicate that the percentage of administrative and compliance costs as a proportion of the total funds allocated, is likely to fall from a (low, mid-range, high estimate) between 12% to 17% for the period 2004-2006, to between 1.43% to 1.91% for the period 2007-2012, c.f. Table 6.
- 439 This would be a considerable accomplishment and would place the proportion of administrative and compliance costs of the PBRF close to the proportions reported for the UK (0.8%) and Hong Kong RAEs (0.07%). It would place the proportion of administrative and compliance costs of the PBRF at one-quarter of the costs of allocating funds using a competitive bidding model.
- 440 However given the many imponderables involved in forecasting costs for 2007 out to 2012, a realistic goal may be to ensure that the administrative and compliance costs of the PBRF for the period 2007-2012 do not exceed 2% of the total funds allocated for that period.
- 441 Taking these matters into consideration and with the object of providing a realistic goal this analysis suggests that:
- a benchmark for the administration of the PBRF and compliance costs for TEOs be set; and
 - that reaching that benchmark be a goal for the 2007-2012 PBRF.
- 442 We recommend

That the administrative and compliance costs of the PBRF and 2006 Quality Evaluation for the MOE, the TEC and TEOs, be no more than 2% of the total PBRF funding allocated for the period 2007 to 2012.

What did TEOs report about their compliance costs in 2003 and their expectations for 2006?

Compliance costs to TEOs from the implementation of the PBRF 2003

443 Cost drivers for all TEOs

TEOs in this evaluation and some TEOs contacted during the evaluation, provided estimates of their compliance costs for 2003. Common costs identified by those TEOs were associated with:

- understanding and clarifying the requirements of them from the beginning of the implementation process in early 2003;
- preparing for and participating in the preparation of policy and guidelines;
- appointing or redirecting staff to part or full-time work over 2003 to manage the TEO's participation in the 2003 Quality Evaluation;
- preparing advice and delivering support to staff to assist them to prepare and check EPs that complied with the *Guidelines*;
- aligning their internal research recording and management systems to enable them to meet the reporting and submission requirements of the PBRF;
- preparing or purchasing software systems to support EP preparation because the TEC software was not expected in time. Some TEOs commissioned their own web-based EP preparation and recording software because their judgement was that the large number of EPs to be prepared, internally assessed and submitted meant they could not wait for or rely upon the TEC supplied platform;
- establishing and carrying out an internal peer review process, including the assignment of Quality Categories, the moderation of the process and the handling of appeals;
- coordinating all activities to ensure EP submission was completed on time;
- liaising directly with the TEC to clarify matters during 2003; and
- liaising with other TEOs.

444 TEOs estimated that of their estimated total compliance costs in 2003:

- these tasks accounted for 20 to 30% of their total costs;
- 5% of their total costs were for IT, library and other services; and
- 65% to 70% of their total costs were associated with staff time.

445 Cost drivers that varied for different TEOs.

The estimates of costs provided by TEOs in this evaluation and some TEOs contacted during the evaluation showed that the following factors influenced the costs to TEOs unevenly:

- the number of staff eligible to participate in the internal processes, larger TEOs had higher costs than smaller TEOs;
- the number of staff eligible to participate in the internal processes not requiring high levels of advice and assistance. A well-informed and experienced staff accustomed to recording their research accomplishments reduced the support required to explain and assist them. TEOs without established research cultures or research support structures spent proportionately more time supporting their staff;
- the ability of some TEOs to collect and lodge ROs, and to a lesser extent CRE and PE, into EPs electronically;
- the amount of material staff had to sort and select for inclusion in their EPs. One TEO completed EPs at very little cost in time to staff because it had an effective IT platform and because it had relatively few PBRF-eligible staff with relatively few ROs;
- appointing a Project Manager to co-ordinate the TEO's response. For some TEOs this was an additional position, for other an existing person altered their responsibilities;
- TEOs having multiple campuses where staff required advice and assistance thus incurring travel and accommodation costs. Multi-campus TEOs reported that they believed that they faced additional costs compared to single site TEOs; and
- the strategy adopted by TEOs with regard to internal peer review panels. Some TEOs established internal peer review panels, some did not; some TEOs employed international independent panellists, most did not; some TEOs made extensive efforts to improve staff EPs, others made more modest efforts.

446 As the PBRF is implemented TEOs are discovering ongoing costs. When TEOs received the results of the 2003 Quality Evaluation and their funding allocations they incurred costs in:

- analysing the results, comparing them with their own assessments;
- reporting the results of the external assessment process to staff;
- investigating, preparing, lodging and pursuing complaints; and
- for some, advertising associated with the results.

447 TEOs reported that their costs were reduced by:

- academic staff and the internal PBRF support team having a sound initial understanding of the requirements of the Quality Evaluation. This enabled some TEOs to adjust more rapidly than others as the TEC *Guidelines* were promulgated; and
- existing resources (designated persons or responsibilities or recording systems) could be integrated with the requirements for compliance with the 2003 Quality Evaluation.

448 During this evaluation TEOs reported that now that they had been through one year of the PBRF and a 2003 Quality Evaluation, they had more grounded understanding of how the many parts of the PBRF worked together. In particular they reported an improved understanding of some actions that might:

- minimise their annual and periodic costs; and
- maximise their annual and periodic return.

How the 2003 baseline experiences of TEOs alter their behaviours in the second allocation period will need to be monitored. It may be that the full range of strategies and new behaviours will emerge in 2012.

449 Some TEOs reported that as part of their management of costs they intended to review:

- their recruitment and appointment strategies – in the light of the employment criteria;
- the employment terms and conditions they may propose to certain categories of staff – in the light of the substantiveness test;
- the nature of employment agreements they may prefer to offer staff – in the light of the employment criteria and the substantiveness test;
- the human resource management strategies they might to take in the matter of staff performance management - in the light of the reporting framework; and



- the alignment and integration of their internal research management information systems, data collection and reporting arrangements with the TEC's IT platform.

The costs of staff time

- 450 The unit of assessment for the 2003 Quality Evaluation was individual academics. PBRF-eligible individuals were required to complete EPs. For some gathering research records was a substantial project, for others it was less so, but for all judgements were required. TEOs provided advice, guidance and support to individuals. The ability of TEOs to provide advice, guidance and support to individuals varied.
- 451 Estimates of the time commitments of individual participants varied. Factors reported in the evaluation as influencing the time required for individuals to complete EPs included:
- their initial understanding of what was required of them;
 - their access to sources of guidance, if they existed, to clarify their understandings of what was required of them;
 - the technical platform (IT) they worked upon. Apple users could not use the TEC software;
 - their PBRF EP readiness. That is whether their NROs, PE and CRE records existed or they required time to gather them; and
 - the accuracy and timeliness of the advice, guidance and support their institution was able to provide.
- 452 Internal peer assessment panel members and HODs provided further time to the process. The internal peer assessment process adopted by a TEOs influenced the time required of:
- internal panel members to read, score, discuss and advise on EPs;
 - HODs to review EPs, advise and manage staff participation in the internal process; and
 - individual staff members to complete EPs, review and submit EPs.
- 453 TEOs adopted a range of internal review processes. Some TEOs established peer preview panels and included international peer reviewers while others did not establish panels. Some TEOs conducted their panels early in the year in order to provide individual advice to all PBRF-eligible staff before EPs were finally submitted to the TEC.
- 454 One TEO in the evaluation believed that the scores allocated by the internal panel should approach the scores allocated by the peer review



panels. The TEO reported that it was considering using the 2003 results to prepare a more exacting internal assessment process in 2006. The TEO was considering inviting the TEC to send a 2003 peer review panel, or part of a panel, to the TEO to work through some example EPs alongside prospective internal panel members. This was to ensure that internal scoring became consistent with peer review panel scoring. Such preparation will increase the time demanded of staff in the 2006 Quality Evaluation.

- 455 TEOs that provided estimates of staff time costs based their estimate on an average time taken by a staff member to complete an EP, multiplying those hours by total PBRF eligible staff, dividing the total to give a number of FTEs and then costing the FTEs at the rate of the annual salary of a senior academic. Some estimates of average time included an allowance for staff time spent on internal panels and HOD time spent on checking EPs and assisting staff.
- 456 TEOs estimated that up to 70% of their total compliance costs in 2003 were associated with the preparation, completion, checking and assessment of EPs by academic staff and HODs and by academic staff preparing for and participating in internal peer review panels.
- 457 TEOs were very clear that staff time demands and the costs of that time were extremely high in 2003. The estimates provided by the nine TEOs in Tables 3,4 and 5 enable an average cost per PBRF-eligible staff member to be calculated. Because of the variations in the methods and assumptions applied by TEOs in making their estimates, and the additional assumptions required to make the estimate, the figures in Table 9 are very tentative indeed.
- 458 The data in Table 9 indicates that, in the TEOs that provided data, the average cost for each PBRF-eligible staff member's participation in the 2003 Quality Evaluation and 2004-2006 PBRF ranged between \$1,400 dollars and \$2,200 dollars.
- 459 No estimate of the average cost for each PBRF-eligible staff member's participation in the 2006 Quality Evaluation and 2007-2012 PBRF is made because those will be virtually identical with the estimates in Table 9. This is because the estimates for 2006 are based on the 2003 estimates with no new data.

Table 9: Estimated average cost per PBRF-eligible staff member for 2003-2006 for nine TEOs.

	(\$)
Low estimate	\$1397
Mid-range estimate	\$1895
High estimate	\$2295

Notes:

- i. The estimates used in Table 9 are based on the estimates in Tables 2 and 3.
 - ii. The cost estimates for the nine TEOs were extrapolated to all participating TEOs. The estimated costs to TEOs' of compliance with the PBRF for the years 2004-2006, see Table 4, were then added to the three (low, mid-range, high) estimates of costs.
 - iii. The estimated total costs to TEOs for the period 2003-2006 (low, mid-range, high) were then divided by the total PBRF-eligible staff, 8017, to provide the estimates above.
- 460 In a recent commentary on the UK RAE, the Times Higher Education Supplement² (THES) estimated the average cost (regulation, accountability and compliance) per research -active staff member spread over six years, as £1,200 pounds (c. NZ \$3,500).
- 461 The THES article indicates that the estimated average costs per PBRF-eligible staff member for the PBRF 2003-2006 at the low, mid-range and high estimates (Table 9), are all lower than what the THES reports are the average costs per research-active staff member in the UK RAE.
- 462 All TEOs in the evaluation were confident that they would be able to reduce the time demanded of staff in the 2006 Quality Evaluation.

The costs of the 2003 Quality Evaluation

- 463 The 2003 Quality Evaluation was the first. The combination of design elements, complexity, newness and demands on staff time ensured that the costs of the implementation of the PBRF and the conduct of the 2003 Quality Evaluation were always going to be high. The costs were also high as a proportion of the total funding allocated because the total funding allocated by the PBRF in the period 2004-2006 was small.

² Times Higher Education Supplement, June 18, 2004 p.6

- 464 The design elements of the PBRF that were associated with costs were:
- conducting both external and internal peer review panels;
 - an audit of (up to) four NROs in each EP conducted on behalf of the TEC by the National Library of New Zealand;
 - an audit of the PE and CRE components of the EPs conducted by the TEC (with assistance from the National Library of New Zealand)
 - an audit of staff eligibility led by the MOE; and
 - an independent assurance over the processes of the Quality Evaluation conducted by the Office of the Controller and Auditor-General (OAG); and
 - using an individual unit of assessment.
- 465 The PBRF Working Group expected that in the future the external panels would cease and there would be internal panels only, a shift to self-assessment and the removal of a substantial periodic cost. The cost of the 2003 Quality Evaluation peer review panels was \$1.1 million dollars; removing those would make savings.
- 466 However this evaluation suggests that TEOs are more likely to prefer to retain peer review panels but hold mixed views about the formal or necessary involvement of internal panels after the 2006 Quality Evaluation. Some TEOs linked their view on retaining internal panels with how effectively eligibility was resolved. They saw internal panels as taking a larger role in determining which EPS were submitted and so reducing the costs of the peer review panels. This is a matter to be monitored and considered in the Phase 2 of the Evaluation Strategy.
- 467 Removing the peer review panels, as planned, would reduce administrative costs substantially but it would also remove a critical confidence mechanism required at this stage in the PBRF's development. A fair and robust assessment process was always required and, on balance, TEOs appeared to believe that that was largely achieved in 2003.
- 468 TEO expect that fairness and robustness can only improve for the next Quality Evaluation, especially if eligibility is resolved. What did emerge from TEO reflections on the 2003 Quality Evaluation was a very strong requirement for a consistent national assessment process; the role the peer review panels played in 2003.
- 469 If the TEC alters the peer review panel composition (more disciplines and more experts, more accommodation), its working process (more cross-referrals) and its working spaces (improved working spaces), then that cost of that assessment process will be greater in 2007.

TEO expectations for their costs in the 2006

- 470 Most participants in this evaluation understood that the administrative and compliance costs of the implementation of the PBRF and the conduct of the 2003 Quality Evaluation would be high. Most participants expect that their costs will decrease because there will be some efficiency gains; though it may be fairer to describe this a strong hope.
- 471 The expectation that the costs of the next Quality Evaluation will not be as high are based on the broad sense that:
- the costs of the design and implementation of the PBRF were one-off costs and the PBRF is now in place;
 - the initial Quality Evaluation was prepared and conducted within a very short timeframe at the same time as the PBRF was implemented. Administrative costs for assembling and managing the TEC capability to carry out both tasks urgently were high, as were the demands on TEO staff managing compliance or participating in the Quality Evaluation;
 - the amount of PBRF funding allocated is planned to increase significantly in the future;
 - while there may be increases in some costs there may also be decreases in other costs associated with the activities of the MOE, the TEC and TEOs in the PBRF; and
 - no-one believes the TEC will not be better prepared or that it would expect the sector to show the same tolerance again.

TEO identified possible cost reductions

- 472 TEOs identified the following improvement to the PBRF as likely to decrease their compliance costs in the 2006 Quality Evaluation:
- the establishment by the TEC, of a cross-platform web-based IT platform for EP preparation, maintenance and submission preferably released to the sector early in 2005;
 - resolution of eligibility issues;
 - early release of the guidelines for the conduct of the 2006 Quality Evaluation by the TEC and no substantive changes to those thereafter. This will enable TEOs to align and integrate their internal information and human resource systems with the requirements of the PBRF. TEOs prefer to receive the 2006 guidelines yesterday; and



- improvements to the assessment processes of the quality evaluation as a result of learnings from the 2003 Quality Evaluation.
- 473 TEOs reported that if these matters were resolved successfully and early, then they expected to reduce their estimated costs for the 2006 Quality Evaluation substantially. The main reductions would be in staff time and the removal of errors in interpretation and application of the *Guidelines*; efficiencies are also anticipated from better planning because TEOs would have longer lead times.
- 474 TEOs reported that they expect to maintain EPs on an annual basis as soon as the platform is settled and aligned with their internal systems. Those TEOs that have not established their own web-based systems reported that this might enable their current estimate of staff time costs to be reduced by more than 40%.
- 475 On the basis of this advice, we have estimated that the total compliance cost to TEOs in the 2006 Quality Evaluation will decrease by at least 20% of the 2003 Quality Evaluation compliance costs.
- 476 However the compliance costs to TEOs from their participation in the 2006 Quality Evaluation may increase as a consequence of:
- increases in inflation and staff costs;
 - internal decisions TEOs take in order to increase their gains from the system that may see them increase expenditures towards the limits of the expected marginal return to compliance cost;
 - alterations to the design of significant elements of the PBRF; and
 - alterations to systems and processes in the conduct of the 2006 Quality Evaluation made by the TEC.

Expectations about the ongoing costs of the PBRF

- 477 In the years between Quality Evaluations there are ongoing compliance costs to TEOs associated with their participation in the PBRF. The ongoing costs will include:
- any adjustments to systems in response to changes in the PBRF made by the TEC;
 - determining PBRF eligibility for any new courses;
 - collection and submission of data on RDC and ERI to MOE;
 - audit of ERI calculations;
 - confirming and checking returns with the PBRF administrator; and

- monitoring and reporting on the impact of the PBRF on staff and the TEO.

478 During the period 2004-2006 each new activity associated with TEO compliance with the PBRF will reveals its costs and impacts as it is experienced for the first time. This will be the case for MOE, the TEC and TEOs until the first whole cycle of the PBRF has been completed in 2006.

Periodic costs of Quality Evaluations to the MOE and the TEC

479 The Quality Evaluation is a periodic measure of research excellence undertaken at the level of individual PBRF-eligible academics. Conducting a Quality Evaluation imposes periodic administrative costs on the MOE and the TEC and periodic compliance costs on TEOs.

480 The MOE and the TEC have experienced the initial Quality Evaluation and are preparing for the next in 2006. There may be improvements and efficiencies to be made. Equally, if there are alterations to the peer review panel process or any other process required these may increase the administrative costs to the TEC. The periodic administrative costs of a Quality Evaluation are incurred before and after the year in which it is conducted.

481 Costs to the TEC from conducting a Quality Evaluation include:

- project management and co-ordination, may include fixed-term contractors to secure specific expertise required;
- consultation with the sector to prepare guidelines for the Quality Evaluation;
- preparation of the *Guidelines*;
- selection, establishment and operation of the peer review panels;
- support for TEOs and peer review panels during the Quality Evaluation;
- provision of an IT platform for EP capture;
- provision of audits of staff eligibility and NROs;
- analysis of results;
- reporting of results;
- conducting an appeals process;
- reviewing the Quality Evaluation and developing it; and

- policy development and liaison with MOE.

Periodic costs of Quality Evaluations to TEOs

482 There were high costs to TEOs associated with their compliance with the 2003 Quality Evaluation.

483 However there will continue to be costs to TEOs associated with their participation in the periodic Quality Evaluations. Those costs will include:

- project management and co-ordination;
- understanding the *Guidelines*;
- applying the *Guidelines* – determining eligibility;
- preparing and advising staff on compliance;
- preparing, collecting and processing EPs;
- monitoring and checking NROs;
- assisting audits of NROs and eligibility;
- carrying out peer review internally;
- liaising with the TEC and other TEOs;
- submitting EPs; and
- supporting staff.

484 Whether the costs of complying with the Quality Evaluation will reduce to an acceptable level for all participating TEOs will be a matter for the TEC to monitor in Phase 2 of the Evaluation Strategy.

485 Most TEOs in this evaluation wish to know the guidelines for the 2006 Quality Evaluation as soon as possible in order to begin some of the tasks above. TEOs plan to reduce some of the identifiable costs to them of the 2006 Quality Evaluation by ensuring that they have collected and verified all the data required for EPs and the NRO audit in the years preceding that Quality Evaluation.

Cost issues for tertiary education organisations developing research cultures

486 Some TEOs that participated in the 2003 Quality Evaluation reported that they did not have an established research culture or resources to

support researchers. Some of those TEOs believed that for them, participation in the 2003 Quality Evaluation and compliance with the PBRF was more difficult and more expensive it was for TEOs with established research cultures and research support structures. Some TEOs believed that their lack of experience and resources disadvantaged them and the support they could expect to offer staff in 2003.

- 487 Some TEOs in the study with established research management and support services reported they were able to respond quickly and effectively to the rapid implementation process in 2003. These TEOs had support structures but they also had a much larger number of PBRF eligible staff to prepare and assess within the same tight time frame as the TEOs without those resources.
- 488 Some TEOs who participated in the 2003 Quality Evaluation are in the early stages of a long-term strategy to develop a research culture based on the traditional academic model of research. These TEOs reported that they intend to continue to develop and invest in their relatively new research cultures. Some of those TEOs entered the 2003 Quality Evaluation in the full expectation that they will never score so poorly again.
- 489 However some of those TEOs would reconsider their investment strategy if:
- they conclude that the definition of what counts as research excludes the work their staff do;
 - the return from their participation is judged to exceed the costs of participation in the medium to long term;
 - that they judge that there is no benefit to be gained from further participation; or
 - there is an alternative funding mechanism that they can access more readily, that may be private funding or another Crown fund.

Creation of asset

- 490 TEOs and the National Library of New Zealand have drawn attention to the possibility that a national web-based database of research outputs, such as may result from a TEC hosted web site, would form a national research resource and asset. Some of the implementation and compliance costs of the PBRF would have built that asset. However, such a database will also require ongoing maintenance, and those are further administrative costs.

Monitoring costs

- 491 The PBRF was implemented in 2003 and the initial Quality Evaluation undertaken, the design has been tested by practice. The PBRF will remain in a transition phase until it is fully implemented in 2006/07.
- 492 By 2007 the sector will have more data on the:
- ongoing administrative and compliance costs of the PBRF; and the
 - administrative and compliance costs of the periodic Quality Evaluations.
- 493 The sector will also begin to observe how TEOs respond to the incentives for gaming the allocation system and whether those strategies have an impact on costs in the system. In order to capture the potential gains from the returns available, TEO spending on compliance activities may increase; thus TEOs may increase their costs to increase their returns as against other participating TEOs.
- 494 In that case the administrative costs to the MOE and the TEC may increase as they seek to neutralise the benefits of gaming. It is too soon to identify any early indicators of this trends emerging. The analysis does point to sensitivities that requiring monitoring and possible early intervention.
- 495 By 2007 the sector may begin to see the impact of the PBRF on the development of research cultures, on researcher and TEO behaviour and on research excellence. These are the benefits intended as the trade-offs for a more expensive funding allocation system than the more arbitrary but potentially less manipulable mechanism of EFTS.
- 496 By 2007 the sector may also be in a better position to judge whether the PBRF is likely to deliver the benefits intended at an acceptable level of administrative and compliance costs as a proportion of total funds allocated.
- 497 In *The 2003 Assessment* the TEC noted:
- Without doubt the exercise was demanding, time-consuming and costly – for participating TEOs, individual researchers, the members of peer review panels, and those charged with the implementation of the PBRF within the TEC (para 256).
- Attention will also be given to ways of reducing the compliance and administrative costs associated with the PBRF (Tertiary Education Commission, 2004a, p86).
- 498 The design and operation of the PBRF will remain a work in progress until 2012. The administrative and compliance costs of implementing the



PBRF and conducting the 2003 Quality Evaluation were high for MOE, the TEC, TEOs and many individuals. Whether the level of costs will increase or decrease depends upon refining existing processes and the cost impact of any new processes.

- 499 This analysis indicates that the large increase in PBRF funding available in the future will reduce the proportion administrative and compliance costs are of total PBRF funding. Whether the actual administrative and compliance costs increase or decrease will require ongoing application of the Efficiency principle.

Options for savings

Early release of guidelines for next Quality Evaluation

- 500 TEOs expressed a strong desire that due consideration be given to revising the *Guidelines* and that the guidelines for 2006 be issued as soon as is possible. TEOs see settled guidelines as:
- removing a significant source of uncertainty. This on two levels. First they wish to begin to prepare for 2006 as soon as is possible. Second, they wish to ensure that there is only one set of guidelines issued for the 2006 Quality Evaluation and that there is almost no possibility of updates or substantive amendments after those guidelines are issued. Their sense of last minute rule changes during the 2006 Quality Evaluation remains a frustration. Their desire to advantage their TEO, or at least not disadvantage their TEO relative to others means they wish the rules, especially around eligibility, declared so that they can action medium and long term strategies around employment agreements and hire dates, from the beginning of 2005; and
 - enabling them to reduce their ongoing costs and improve the quality of their investment decisions because they expect to have time to align their research and information recording and management systems, (IT and people systems) with the settled PBRF requirements.
- 501 TEOs will see a TEC commitment to the speedy release of revised guidelines as the first indicator as to how the 2006 round might proceed. TEOs in the evaluation agreed that despite the frustrations, the TEC had been open, used its best endeavours and good people, and in the end, done pretty well to achieve what it did.
- 502 However there is a strong view that the timetable of 2003 cannot be repeated. TEOs will observe the TEC's actions to see if the review, planning and preparations for 2006 are planned and begin very soon. If they do not see that that is the case they may conclude that the TEC

plans or may be inadvertently placed to repeat the preparation and timeline of 2003; that would prompt anxiety in the sector.

- 503 For this reason we have emphasised that retaining the sector's goodwill and constructive engagement from the 2003 Quality Evaluation is extremely important for the TEC. It may also help to reduce the costs to TEOs of their preparation and compliance with the 2006 Quality Evaluation.
- 504 This report emphasises that the PBRF and the Quality Evaluation process have had only their very first round, they are not settled, they require development and refinement and their impacts are yet to emerge. This report has also emphasised that if the 2006 Quality Evaluation is to proceed as well as participants in this evaluation wish it would, then preparations should be seen to be underway.

Undertaking to have web-based EP solution in place by the end of 2005

- 505 TEOs in the evaluation were strongly of the view that a sector-wide web-based EP preparation, maintenance and submission platform would reduce substantially the costs of their participation in the next Quality Evaluation. TEOs see a purpose-designed EP platform as:
- enabling them to establish mirror sites within their own organisation that they can maintain in the periods between Quality Evaluations;
 - ensuring compliance and consistency in EP format and filling;
 - providing a research asset; and as
 - reducing the risk of error at the time of submission of EPs to the TEC.
- 506 Some TEOs have developed internal IT platforms for EP management while others are waiting for an indication by the TEC whether it intends to commission and issue a web-based IT platform for the preparation and management of EPs, across operating system platforms, in time for the 2006 Quality Evaluation. Whether the TEC intends to seek funding to commission the web-based IT platform TEOs can imagine is a matter for the TEC; no promises can be made in this evaluation.
- 507 The possibility of integrating PBRF compliance instruments with the ongoing operations of TEOs was one mechanism that *Investing in Excellence* indicated would offer the possibility for reducing the administrative and compliance costs of the PBRF.

The understanding was that many providers were already implementing Research Assessment Exercises or Institutional

Grant Scheme-style systems for internal research grant allocations (Ministry of Education and Transition Tertiary Education Commission, 2002, p10).

- 508 We have recommended that the TEC commission such a platform. To be in place for the 2006 Quality Evaluation, the platform would need to be designed, built and implemented before late 2005. This would require an early TEC decision and substantial (\$1.5 to \$2 million) additional funding to the TEC.

Cessation of internal peer review after the next quality evaluation

- 509 *Investing in Excellence* considered that, in time, the peer review panels could be removed from the Quality Evaluation process. This would leave the internal TEO peer review panels to undertake the assessment of the EPs of PBRF-eligible staff. Following the 2003 Quality Evaluation, the removal of the peer review panels is most unlikely for 2006.
- 510 The peer review panels were a significant cost (in time, money and energy) of the 2003 Quality Evaluation. Many panel members we observed, in particular panel chairs and international reviewers, committed long hours to reading *Guidelines*, EPs and attending meetings. Some panel members interviewed explained that they did so because:
- they wished to understand their task thoroughly before they began to assess EPs; and because
 - they wished to undertake their assessment with as much integrity as they could so they undertook more tasks, i.e. reviewed more EPs, read more NROs, than was required.
- 511 Some panel leaders we observed went to considerable lengths to ensure that their panel made a fair and thorough assessment of a large number of EPs. This demanded a great deal of their time and appeared to be an unsustainable expectation in small academic community; 'Who would do that again?' some wondered.
- 512 As this report noted earlier, some TEOs in the evaluation are considering making their internal panel process more thorough. This may increase the time required of staff when efficiencies and improvements in the overall process were intended to decrease the time demands on staff.
- 513 If the costs of the peer review panels remain approximately the same or, as is more likely the case, are higher for the 2006 Quality Evaluation because of alterations to the assessment process, then there may be merit in removing the requirement for internal panels after the 2013 Quality Assessment.

Reducing administrative costs at the center

- 514 There appear to be limited opportunities for making savings in the administrative costs of the PBRF in the transition period 2004 to 2006. The costs to MOE are projected to diminish but the TEC has little time to review the 2003 Quality Evaluation, undertake any redesign and prepare for the 2006 Quality Evaluation.
- 515 After 18 years, the central administration of the UK RAE is small and practiced. The RAE Manager's Report on the 2001 RAE discusses:
- the planning and development cycle for the 2001 Exercise undertaken following the 1996 RAE; and
 - the Project Management, Project Control and Budget for the 2001 RAE exercise.
- 516 To prepare and conduct the 2001 UK RAE a manager was appointed in November 1998, an initial team of five persons was appointed and that grew to thirteen during 2001 in order to conduct the RAE. The project manager reported to a Steering Group appointed by a Project Group formed by the funding bodies. This team worked closely with colleagues in the Analytical Services Group of the HEFCE who managed the data collection. Following a tender process the RAE developed the 2001 Data Application software that was provided to HEIs to enable them to submit data.
- 517 The RAE Manager's Report on the 2001 RAE notes that:
- the initial operating budget for the 2001 RAE had been set at £3,602,983 in 1997 but that the actual cost was £5.109 million. The actual costs of the 2001 RAE compared to the estimated cost of the 1996 RAE of £3.032, was an increase of 68%; and
 - many of the cost increases related to improvements in the RAE process since 1996, including the participation of users, an increased number of panel meetings, greater level of scrutiny of submissions by panels, more extensive use of specialist advisers (and paying them for the first time in 2001) and of cross referral, the establishment of large sub-panels in medicine, mandatory cross-referral for interdisciplinary research and the introduction of Non-UK-Based Advisers (Rogers, 2002, pp18-19).
- 518 The cost of the 2001 RAE above includes the administrative costs at the centre for the period since the 1996 RAE but not the costs of compliance for HEIs during that period or during the 2001 RAE.
- 519 The 1999 RAE in Hong Kong was undertaken by an Assistant Secretary Responsible for Research Administration, an executive assistant and two temporary clerical staff, four people (Boston, 2002, p84). The PBRF is

unlikely to ever approach that level of efficiency if only because employment regulations protect employees against excessive stress.

- 520 The TEC Board established a Steering Group to govern the project and contracted a Project Manager. The PBRF team began as approximately 10 FTEs in early 2003 and grew to approximately 20 FTEs for the full implementation period. The PBRF implementation team had a much greater task in 2003 than either UK or Hong Kong RAE teams had in running their established RAEs in 1999 (Hong Kong) or 2001 (UK).
- 521 The PBRF Project Manager's report comments that the flat management structure of the PBRF Project and the delegated authority of the Steering Group provided an effective governance and management model.

The Project Manager for the PBRF implementation attended Steering Group meeting and acted as the secretary to the Steering Group. The effectiveness of this structure can be gauged from the level of acceptance of the project within the sector, the robustness of the processes, and the quality of the reports. From a project management perspective the pragmatism and willingness to delegate important decisions when time was on the essence on the part of the Board, the Steering Group and the Managing Director were instrumental in the success of the exercise (Tertiary Education Advisory Commission, 2003, p3).

- 522 The PBRF Project Team was able to make decisions quickly and respond to the sector's needs during the implementation process; this rapid decision making also assisted with the success of the sector liaison role.

An essential attribute of small teams undertaking complex projects with many components that vary over the life of the project is flexibility. This was achieved through recruiting for a high level of competence, ensuring that all staff are informed about and familiar with all aspects of the project and in maintaining a flat management structure with the shortest possible reporting lines (Tertiary Education Advisory Commission, 2003, p4).

- 523 To undertake the simultaneous implementation of the PBRF scheme and prepare and conduct the initial Quality Evaluation was a major project on any measure. The TEC elected to use fixed term contracts to secure many of the skills and people it required for the tasks. The Project Manager reported that:

Heavy reliance on contract consultants has proved a mixed blessing in the context of the 2003 round. An advantage has been that a very high quality of staff has been secured when required.....

Thorough documentation, decision recording, and filing militate against the often-voiced need for continuity arising from the knowledge loss with staff departing. Additionally a significant number of permanent TEC staff have a detailed knowledge of the processes including the Managing Director, the steering group members and Policy and Communications staff (Tertiary Education Advisory Commission, 2003, p3).

- 524 Immediately after the completion of the 2003 Quality Evaluation, the implementation team was disbanded and a small team of approximately 3 FTEs retained to manage the evaluation and appeals process, wrap-up the 2003 Quality Evaluation and begin preparations for the 2006 Quality Evaluation.
- 525 The PBRF has been implemented and will be in a transition phase until the next Quality Evaluation. It is important that the ongoing administration of the PBRF goes well in the period 2004-2006, but it is particularly important that the 2006 Quality Evaluation go well.
- 526 This report has emphasised:
- TEO expectations that preparations for the 2006 Quality Evaluation will begin shortly;
 - TEOs confidence in the 2006 Quality Evaluation will be maintained if TEOs see and experience the TEC as operating a longer planning cycle for the next Quality Evaluation than for the 2003 Quality Evaluation; and that
 - TEOs in the study reported that they would be very unhappy if the knowledge and expertise about conducting the PBRF and Quality Evaluation process acquired by National Library of New Zealand Staff, MOE and the TEC appear not to have been captured when preparations for 2006 begin 2006.
- 527 This report suggests that the TEC consider whether a small, but visible (to the sector) PBRF team should be kept in place to prepare for the 2006 Quality Evaluation. This is in contrast to the option of contracting in an implementation team that would be largely responsible for the 2006 Quality Evaluation, late in 2005 or early 2006.
- 528 For these reasons it may be only after the 2006 Quality Evaluation that the TEC can consider a more efficient governance and management model for the PBRF.

Chapter 5: Assessing Evidence Portfolios

Introduction

- 529 This chapter considers the Quality Evaluation process. Design decisions about eligibility and the unit of assessment are considered separately. Where a process matter is a consequence of one of those issues it is discussed in this chapter.
- 530 One component of the evidence collected for this evaluation was an on-line survey of chairs and members of the peer review panels. The survey had an 83.3% response rate. Evidence from the survey is used in this chapter. Appendix 2 is a report on the survey.
- 531 The assessment of individual staff members' EPs by the peer review panels was the central element of the 2003 Quality Evaluation. Assessing the quality of research is inescapably subjective. To manage that subjectivity standards of quality and standards of evidence were defined and made known to panel members.
- 532 The panel members were required to exercise their judgements against those standards so that individuals perceive the results as having been made in a robust and consistent manner and so accept their allocated Quality Category as fair in process and in outcome. Further defences against inconsistency were built into the process in the form of multiple levels of moderation at the individual, panel and inter-panel levels. This evaluation concentrates on how well the 2003 Quality Evaluation managed the assessment by the peer review panels.
- 533 Fairness, robustness and consistency are not themselves objective and value free concepts. Those involved in quality assessment in any capacity will have different views as to the meaning of those concepts. A central aspect of this concerns personal views about what constitutes greater or lesser quality. It is not surprising that the evidence gathered for this evaluation showed this to be the case for the 2003 Quality Evaluation.
- 534 The design of the assessment process sought to ensure that as far as possible, those conducting the assessments:
- had a common understanding of the standards to be applied;
 - were able to apply them consistently and fairly; and
 - had a common understanding of how to value different kinds of evidence.

There was a rigorous moderation process that sought to identify errors or inconsistencies and correct for variance in assessments.

- 535 The factors that determined the levels of fairness, robustness and consistency of the quality assessment were:
- the consistency, clarity and completeness of the evidence presented in the EPs;
 - the interpretation of the definition of research and its application to the evidence presented;
 - how well prepared members of panels were for their task;
 - the clarity and defensibility of the defined standards and their relationship with the scoring system;
 - the clarity, defensibility and precision of the descriptors of quality and *Guidelines* as to how the evidence was to be assessed;
 - the extent to which individual peer review panel members managed their own subjectivity;
 - the effectiveness of the checks and balances built in to the panel processes; and
 - the effectiveness of the moderation procedures.
- 536 This evaluation does not consider whether these mechanisms and processes can be made to produce a perfectly fair, robust and consistent assessment of research quality that satisfies everybody; that is impossible. Instead the evaluation sought to understand:
- whether the assessment processes designed for the 2003 Quality Evaluation were fit for their purpose;
 - whether there are indicators that the results of the 2003 Quality Evaluation were within sufficiently narrow margins of error and distortion to produce a funding distribution to TEOs and individual Quality Categories, that were perceived to be reasonably fair, robust and consistent; and
 - whether it is possible to improve the existing processes so that they produce results in future Quality Evaluations that reduce error and distortion, and are perceived as more fair, robust and consistent.
- 537 The evidence drawn on for this chapter was obtained from the following sources:
- observations of four peer review panel meetings and all the Moderation Panel meetings;
 - the data from a survey of peer review panel chairs and members;



- interview data with a range of individuals and institutions across the tertiary sector between October 2003 and March 2004;
 - in-depth interviews with staff from the TEC, in case studies at three TEOs, and in initial scoping studies at two other TEOs;
 - analysis of the TEC's internal and external reports;
 - submissions to the TEC and to WEB Research;
 - other documents; and
 - commentary on emerging impacts following the April 2004 release of the results.
- 538 The evidence from these sources suggested that there were few in the sector or the TEC who expected a completely satisfactory outcome from the 2003 Quality Evaluation and PBRF round. Participants generally recognised that the TEC and the sector were engaged in a first implementation that would provide a basis of experience to inform improvements in future Quality Evaluations.
- 539 The on-line survey indicated that a very high percentage (92.9 %) of panel members surveyed agreed that, overall, the peer review panel process was sound. A high percentage (88.6 %) also agreed that the composition of the panels ensured that they made balanced judgements. Some of the open-ended comments revealed unease over possible conflicts of interest and the skills and/or knowledge of the panel members involved.
- 540 While panel chairs and members gave positive responses overall to questions of validity and reliability, there were more mixed responses in the observational, interview, and secondary data, and the qualitative survey data.
- 541 These overall differences can be attributed, in part, to the nature of the on-line survey population. The population consisted of panel chairs and members who made a substantial time commitment to the process. They certainly worked exceptionally hard to make their part of the assessment process work.
- 542 No expectation of significant changes in the design of the mechanisms and processes used emerged from our observations. Participants expected that modifications would be made to improve the fairness, robustness and consistency of future Quality Evaluations.
- 543 The evaluation evidence suggests:
- that there were some significant distortions in the results of the 2003 Quality Evaluation;

- that fundamentally redesigning assessment processes is unlikely to mitigate these distortions any better than modifying existing processes;
 - there is an expectation in the sector that the 2006 Quality Evaluation process will be modified, but not be substantially different from that of the 2003 Quality Evaluation; and that some planning based on that assumption has already taken place;
 - that there are significant opportunities for improving the fairness, robustness and consistency of the Quality Evaluation process by making various modifications to it; and
 - that (based on the experience of other jurisdictions) it is highly unlikely that any approach to the assessment of the quality of research will ever be entirely fair, robust and consistent.
- 544 Notwithstanding this, there remains some ambiguity about the policy intent of the PBRF. A recurring theme in the data is whether the emphasis in the PBRF was to reward and attempt to foster improvement in all researchers, in all degree-awarding TEOs in New Zealand, or 'to concentrate the research funds where it would have most impact in fostering leading-edge research.'
- 545 The undertaking and evaluation of future Quality Evaluations may indicate a need for fundamental redesign, but it is too soon to suggest significant redesign on the basis of this evaluation.

The status of the findings

- 546 The recommendations in this report about the peer review panel process are formed from the data sources listed in paragraph 8. Analysis of this data illuminated some key design and implementation issues and their possible implications. Except in certain specified cases, however, inferences cannot be drawn about the extent of the patterns we describe. That is a matter for the Phase 2 Evaluation Strategy.

The structure and composition of the peer review panels

- 547 A TEC internal report concludes that:

overall, the nomination, appointment of panel members and convening of panels was a success. In less than three months, 580 nominations were received, and 154 panel members and 12 panel chairs were appointed to the TEC PBRF peer review panels. Timeframes were met, appropriate panels were appointed and sector feedback has generally been positive (Tertiary Education Commission, 2004b, p1).

- 548 Nonetheless, the report notes a number of areas where improvements can be made in future. This evaluation endorses such suggested improvements.
- 549 The fundamental decision to adopt a small number of multidisciplinary panels (like Hong Kong) rather than a larger number of disciplinary panels (like the United Kingdom) was considered to be inevitable for a small country. The evidence of this evaluation is that this produced some apparent inconsistencies in assessments of EPs in subject areas where there was no person from the same subject area on the panel, or where the same disciplinary area was assessed across more than one panel. However, this evidence is not conclusive and closer study would be required to verify this inference.
- 550 There are two factors to consider with respect to the peer review panels:
- the number of panels and the subject areas they were assigned; and
 - the representativeness of panel membership.
- 551 The number of panels was broadly a function of the decision to use multi-disciplinary panels. We see no immediate imperative to change that decision, and there are considerable potential problems in doing so, not least questions of logistics and costs. The exact number of panels is not in itself significant and could be changed. However, we believe that the best approach after one round is to keep the number of panels stable, but to adjust the assignment of subjects to each one to achieve a better workload balance if that is possible. There are a number of opportunities to adjust the assignment of subject areas to panels.
- 552 The question of the representativeness of panel membership relates to the proportion of members of a panel who were specialists in each of the disciplines assigned to it.
- 553 The balance of membership of panels created some problems, particularly with respect to:
- workload imbalances between panel members; and
 - the subject area membership of panels not balancing the proportion of EPs in different subject areas allocated to panels.
- 554 According to our interview evidence from those in the TEC and the MOE who were involved in the planning and appointment of panels these issues were primarily a consequence of:
- the hurried timeline for making appointments;

- the timing of the appointments process, which occurred during the holiday period; and
- assumptions about the balances of disciplinary and sub-disciplinary specialist requirements that were based on course and teaching information more than on information about areas of research.

555 The 2003 assessment has now provided much better information about the profile of research being undertaken in TEOs in New Zealand, and especially about interdisciplinary research. A close analysis of the 2003 EPs and panel membership, together with consideration of submissions from TEOs during the Phase Two evaluation, would allow better judgements to be made about panel membership in the future.

556 The evidence from this evaluation also raises the question of why external audits were carried out on a number of aspects of the PBRF process, but not on the panel selection and appointment process. For example, we note that chairs appointed panel members (*ibid.*, p.12). Analysis of the proportion of panel members from the home university of the chair showed marked variation among the panels. An external audit could have provided assurance of the fairness of this selection.

The phases of the 2003 Quality Evaluation

557 The 2003 Quality Evaluation process had seven main phases:

- EPs were compiled within TEOs, they were assessed internally and then submitted to the TEC;
- the TEC received EPs, validated the data and checked the EPs against the panel alignment;
- the TEC assigned the EPs to panel members and the panel chair gave approval;
- EPs were assessed by panel members and these assessments were analysed by TEC staff and discussed by the Moderation Panel. Feedback from the analyses was sent to the panels;
- the panels met and assessed the EPs;
- the panels gave feedback to the Moderation Panel and checking processes took place where necessary;
- the panels recommended Quality Categories to the TEC Board.

558 The Assurance Audit conducted by the Office of the Controller and Auditor General found that:

nothing has come to our attention that causes us to believe that TEC's processes, procedures and practices, used during the evaluation of evidence portfolios submitted by Tertiary Education Organisations, were not conducted fairly and objectively (Tertiary Education Commission, 2004a, p264).

- 559 The Office of the Controller and Auditor General concluded that the overall process was observed to be robust and conducted in terms of the *Guidelines*.

The preparation, assessment and submission of EPs in TEOs

The implementation of the phase

- 560 The EPs and the evidence presented in them were the key component of the Quality Evaluation. Every PBRF-eligible staff member completed an EP, providing evidence of three components:
- Research Outputs (ROs). These were the outputs of a staff member's research. Up to four could be nominated (NROs), which were to represent the staff member's best work. NROs had to be available for examination by peer review panel members. Up to 50 other ROs could be included. All ROs had to meet the criteria set out in the *Guidelines*, and to have been produced in the period 1 January 1997 to 31 December 2002;
 - Peer Esteem (PE). This was an indication of the quality of the staff member's research as recognised by peers; and
 - Contribution to Research Environment (CRE). This was an indication of the staff member's contribution to the research environment, both within the TEO and beyond it.
- 561 The *Guidelines* required that the information used by the peer review panels in their assessments consisted only of what was contained in the EPs prepared by eligible staff, examination of a proportion of NROs, together with the results of internal quality assessments carried out within TEOs.
- 562 TEOs were responsible for:
- determining the eligibility of staff to participate;
 - ensuring that eligible staff completed EPs; and
 - submitting EPs to the TEC.



- 563 The Draft Guidelines (Quality Evaluation – Completion and Submission of an Evidence Portfolio) were detailed in the Strategic Consultation Draft Guidelines posted on the TEC web site on 31 May 2003.
- 564 The Draft Guidelines for the 2003 Quality Evaluation were consolidated and issued to TEOs on 25 July 2003 as the '*Performance-Based Research Fund: A Guide for 2003*' (*Guidelines*). These *Guidelines* altered the advice to TEOs on concerning the preparation and submission of EPs to the TEC in the following ways:
- the addition of a new section on Pacific Research;
 - the addition of material on forms and formats of NROs selected for examination; and
 - a small number of typographical, stylistic and grammatical amendments.
- 565 Most TEOs, and all of the larger ones, dedicated administrative and management time to assisting eligible staff in preparing their EPs and in conducting the internal assessment.
- 566 The TEOs and staff commenced compiling and assembling EPs following the release of the March draft Guidelines but before the release of the final *Guidelines* on 25 July. During this period the staff managing the compliance process within TEOs drew heavily on:
- advice and guidance provided by TEC staff; and
 - an informal but well organized network of project managers in TEOs.
- These sources of advice and guidance continued to be important after 25 July, but the definitive reference point became the (consolidated) *Guidelines*.
- 567 The TEOs were required to nominate a peer review panel for each EP using the information provided in the *Guidelines* concerning the subject groupings for each of the peer review panels. The subject areas used for both the allocation to peer review panels and for grouping for the purpose of reporting could not always be that represented by the staff members' academic departments.
- 568 The TEOs submitted EPs to the TEC by 30 September 2003. In practice the vast majority of EPs were submitted very close to the deadline.



Fitness for purpose of the phase

- 569 A high percentage (90.0 %) of on-line survey respondents agreed that the EPs provided a valid measure of individual research performance³. While there was a high level of agreement that ROs are valid indicators of research quality (92.8 %), lower levels of agreement were noted regarding PE (79.3 %) and CRE (71.4 %).
- 570 Preparing EPs was largely the responsibility of individuals, and TEOs gave variable assistance. However this was the first experience of the process and it is not surprising that the preparation of EPs was not done consistently well. As a consequence the overall results are likely to be poorer than might have been expected.
- 571 All the TEOs that chose to participate in the 2003 Quality Evaluation managed to comply with the requirements despite the highly compressed timeline.
- 572 The evidence we gathered indicates that some TEOs faced abnormal difficulties with the process where:
- a TEO had multiple, geographically distributed campuses that made it difficult to adequately assist staff and manage the process;
 - a TEO did not have a well developed research culture, resulting in some staff having additional difficulty in understanding what was required in order to comply; or
 - a department or subject area did not have a well developed research culture, resulting in some staff having additional difficulty in understanding what was required in order to comply.
- 573 The existing requirements may present continuing administrative difficulties for TEOs with multiple campuses, but in the cases of problems associated with weakly developed research cultures we were invariably informed that the institution or department intended to learn from the 2003 experience and, in the words of one informant, '*be ready next time.*'
- 574 Every participating TEO complied with the requirement to undertake an internal assessment of EPs. Most, but not all, sought to model the peer review process used by the TEC.
- 575 The TEOs did not submit EPs they considered to be an 'R' category to the peer review panels. This was the main reason why of 8,018 eligible staff, 5,776 had their EPs assessed by a peer review panel, c.f. Table A-1 (Tertiary Education Commission, 2004a, p96).

³ The response rate for the survey was 83.3%

576 We conclude from our case study evidence and the high level of successful compliance that the design of this phase was fit for purpose and requires no substantial change.

Errors and distortions in this phase

577 The data gathered from case studies and peer review panels makes it clear that internal assessments by TEOs resulted in the overall allocation of higher Quality Categories than were allocated by the peer review panels.

578 There were a few cases where the TEO assessment awarded a lower score than did the peer review panel.

579 In one case, some members of a department in a TEO believed that an EP, which had not been submitted to a peer review panel because it had been awarded an 'R' category in the internal assessment, should have been submitted because, in their view, it would have been awarded a 'C' by the peer review panel. As the evaluation fieldwork was completed before the results were published, we were unable to examine the case or establish the number of such cases.

580 The TEOs made considerable efforts to assess EPs within the *Guidelines*. In doing this they drew on advice and guidance from TEC staff. All participants interviewed in the cases studied affirmed the quality of the support and advice they received from TEC staff.

581 However the performance of TEOs was impaired by a number of factors beyond their control:

- the compressed timeline for the 2003 Quality Evaluation created administrative pressures;
- the timeline pressures, together with the lack of experience of the process, did not permit adequate time for the 'negotiation' of shared understandings of all the key rules between the TEC and TEOs prior to the start of the 2003 Quality Evaluation; and
- in order to meet deadlines some or most TEOs started by relying on draft guidelines that were subsequently changed by the TEC.

582 In some TEOs and some disciplines the absence of an established research culture made it difficult for academics to know what they were required to do in order to comply adequately. This was especially so amongst non university TEOs and amongst Māori researchers. This led the Māori Knowledge and Development panel (MKD) to recommend to the TEC that it take special steps to help Māori researchers to better understand what was required.



- 583 Two factors influenced the quality of the EPs submitted for assessment:
- understanding at the TEO level about what information was relevant and required to be inserted into the templates; and
 - understanding at the TEO level about how that relevant information was to be inserted into the template.
- 584 The first of these factors was influenced by the comprehensiveness and timeliness of advice from the TEC. Reports from TEOs and individuals suggested that the TEC advice was comprehensive but too complex; the advice needed to be broken down into useful templates by TEOs and forwarded to individuals. Once that was done the *Guidelines* were updated and TEOs found updating their internal guides difficult at the late stage. In preparing for the next Quality Evaluation the TEC may wish to review some of the internal guidelines prepared by TEOs for examples of the level and layout of guides that are effective for staff.
- 585 The second of these factors reflected some of the incompatibility issues with the software platform, especially where the TEO had elected not to use the TEC supplied software.
- 586 Smaller TEOs and some larger TEOs that were able to do so received and reviewed EPs and offered individuals advice on how to improve them. Others were not able to do so because of one or more of these factors:
- inadequate internal resources;
 - the difficulty of contacting staff who were geographically distributed over many campuses; and because
 - some staff did not understand the significance of what was required of them and did not make themselves available.
- 587 We noted evidence suggesting that there were also cases in TEOs with established research cultures where the process was taken too lightly by individuals and departments.
- 588 The consequence of all these factors was that not all EPs placed before the peer review panels presented the quality of the research of many individuals as well as they could have. This happened often enough to stimulate every peer review panel report to mention it. In the panels we observed there were frequent expressions of frustration about the variable standard of EP preparation.
- 589 Our interview data indicates that there were a small number of cases where it was not clear which peer review panel to assign an EP to. These cases were primarily concerned with multi-disciplinary research. These cases were regarded as examples of lack of experience with the process

but, as we discuss later, experience alone is unlikely to eliminate such problems.

- 590 The tendency for TEOs to award higher Quality Categories than did the peer review panels had no distorting effect on the final results. Indeed it provided an additional opportunity for checking and standardising the assessments because most panels paid special attention to EPs where there was a difference of two Quality Categories between the category awarded by the TEO internal assessment and that awarded by the peer review panel. Some panels also paid special attention to EPs where they initially awarded a Quality Category that was higher than had been awarded in the internal assessment.

Improving the processes in this phase

- 591 Boston (Boston, 2004) has observed that the original intention of moving to a reliance on internal assessment now seems ill-advised. This evaluation supports that view. The evidence gathered for this evaluation also indicates that conducting an internal assessment in 2003 was an important experience for TEOs. Conducting an internal assessment:

- helped to build understanding of the Quality Evaluation and the standards;
- helped to focus staff attention on the PBRF and what was required of them; and
- assisted in the process by which peer review panels achieved better consistency in the award of Quality Categories.

- 592 This suggests that it may be helpful to retain the internal assessment phase for one more Quality Evaluation but as a developmental activity rather than as preparation for shifting to a method based on internal assessment. The counter view holds that if internal assessment is no longer considered to be an option then the costs and efforts should not be replicated in 2006 merely to build experience.

- 593 There can be confidence that the preparation of EPs will be much better for the 2006 Quality Evaluation because all involved have had experience or will have access to others who have had experience. There will be a need to pay special attention to subject areas, groups or institutions where there are weakly developed research cultures, and to ensure they receive adequate advice and support. This is a responsibility of both the TEC and TEOs.

- 594 Notwithstanding the advantage of experience the high quality advice and support provided by TEC staff during the 2003 assessment needs to be replicated and available for the preparation and conduct of the 2006 Quality Evaluation.

- 595 There were many disturbances due to the compressed timeline. We expect that there should be fewer for the 2006 Quality Evaluation.
- 596 The evaluation suggests that most TEOs did their best in difficult circumstances. Learning from their experiences will enable them to be more effective in 2006 provided that the assessment process remains broadly the same as it was for 2003. Allowing TEOs and their staff to learn from experience is an important consideration when deciding whether or not to initiate significant design changes in the Quality Evaluation.

The TEC reception and validation of EPs

The implementation of the phase

- 597 When the EPs were received by the TEC they were validated by an audit process. The process was amended a number of times during the design phase. The reasons for this are outlined in detail in Appendix C, paragraphs 5-8 of *The 2003 Assessment* (Tertiary Education Commission, 2004a, p1). This section summarises the process used.
- 598 The audit methodology rendered NROs ineligible if:
- they were produced outside the assessment period;
 - they were not authored by the person who submitted the relevant EP; or
 - there was no evidence to confirm their existence;
- 599 In addition the audit was intended to identify 'serious errors', defined as errors of a kind that could materially effect a panel assessor's judgement about the quality of an NRO.
- 600 The audits were carried out by the National Library of New Zealand (NLNZ) on behalf of the TEC and seven audit teams visited TEOs during October 2003.
- 601 In practice, and as is described in detail in Appendix C of *The 2003 Assessment*, timeline and workload pressures resulted in the processes not being carried out as originally envisaged. These problems led to a decision to concentrate on the 'fundamental errors' that were deemed to render an NRO ineligible.
- 602 At the conclusion of the audit team visits 1,446 (6.49% of the total) had been identified as ineligible by the NLNZ audit teams. However follow up work by TEOs reduced this to a final 162 (0.72%) that were confirmed as being ineligible.

- 603 The PE and CRE components of EPs were subjected to a sampling audit of 49 EPs. This is described in paragraphs 22-28 of Appendix C of *The 2003 Assessment*. This did not provide a proper audit, but it did provide information about the difficulties involved in verifying such evidence that will contribute to the development of more robust checking methods in the future.

Fitness for purpose of the phase

- 604 The TEC found itself obliged to redesign the process for verifying the evidence contained in EPs a number of times in the face of the magnitude of the task that the original plan generated in practice. Observational and interview data in the TEC revealed very high workloads and particularly high levels of stress in the period that these audits were being carried out. The problems were exacerbated by:
- the tight timeline;
 - the unexpectedly high workload; and
 - the lack of experience in completing EPs in TEOs and amongst eligible staff.
- 605 The evaluation suggests that, while the problems of checking the evidence in EPs will continue to impose a high workload, the problems are likely to be mitigated by:
- a less pressured timeline;
 - improved processes; and
 - greater experience in completing EPs.
- 606 At present there are good evidence-based reasons to place trust in the professional integrity of TEOs and individual staff. Fairness demands that NROs continue to be checked as comprehensively as possible, but the evidence from the 2003 audit is that very few, if any, staff deliberately or negligently misrepresented the evidence.
- 607 Our interview data supports the view that there were many errors because of misunderstandings of what was required, of inexperience and of haste.
- 608 The evidence from this evaluation is that the auditing process needs to be refined and that there remains sector trust that the audit process and its outcomes will improve. There are opportunities for the TEC and TEOs to address the issues associated with the 2003 audit of NROs in the proposed review.

Errors and distortions in this phase

- 609 There were errors and distortions in the EPs presented in the 2003 Quality Evaluation caused by uncorrected serious errors (and other, less serious, errors). The evidence for this is:
- the identification of many serious errors that were not further checked in the auditing of the contents of EPs;
 - the unprompted awareness of errors amongst peer review panel members; and
 - the general expression of concern about the variable quality of EPs.
- 610 There is no evidence that the serious errors identified inflated the results. Our observations indicated that the errors noted by panel members were more frequently errors that did an individual harm than benefited them. Staff did themselves least justice by leaving their contribution to a multi-authored paper unclear.
- 611 Our observations did not suggest any rationale for adding new layers of cost or complexity to the Quality Evaluation process on the basis of the 2003 experience. Future experience might produce evidence and justification for a more rigorous auditing process.

Improving the processes in this phase

- 612 Evidence from interviews with project managers, liaison personnel, research managers and eligible staff members in TEO case study sites, together with observation of and interviews with TEC staff while they were engaged in managing the verification process, indicate that the suggestions made in the report of the NLNZ provide a good framework for considering the possibilities for improving the processes for auditing the evidence in EPs (National Library of New Zealand, 2003).
- 613 The NLNZ's suggestions are summarised on page 14 of its report. These consist of five key lessons learned, which were:
- there is a need for better bibliographic citations than were provided in 2003. This could be achieved by enlisting the assistance of TEO library staff when EPs are being created;
 - those preparing EPs need to have a better understanding of the audit process than they had in 2002;
 - there is a need for clear understanding of what would constitute evidence of an NRO where part of the item is provided in physical form;

- there is a need for a better understanding of what would constitute independent evidence where the item itself is not provided in physical form; and
 - the software used to generate EPs needs to be reviewed and significantly altered.
- 614 The report also offers five suggestions to the TEC. These were that the TEC seeks to develop a verification method that:
- allows for the creation of bibliographic citations as ROs;
 - incorporates the checking of bibliographic citations at the TEO level;
 - requires academic or library staff to provide independent evidence of the ROs in the Evidence Portfolio;
 - creates a database of ROs from which data can be pulled to create EPs; and
 - uses citation software such as Endnote when creating bibliographic citations (National Library of New Zealand, 2003, p14).
- 615 However by themselves these suggestions may create unduly burdensome administrative requirements for a process that was already seen as excessively burdensome in 2003.
- 616 Another possibility presents itself that could usefully be laid alongside the NLNZ suggestions. We note that at least one TEO has already implemented a process whereby the routine management of the recording of research outputs in the TEO has been integrated with the PBRF evidence requirements for EPs. Such systems would enable the collection of evidence for EPs to occur continuously throughout the assessment period for future Quality Evaluations, obviating the necessity for intensive EP preparation and verification in the year of the assessment. However if TEOs are to be encouraged to establish such systems they would need to be confident that the individual unit of assessment was to be retained.

The assignment of EPs to peer review panels

The implementation of the phase

- 617 The TEOs were required to nominate peer review panels for each EP submitted. Upon receipt of EPs the TEC secretariat checked against the *Guidelines* that the panel nominated covered the subject area identified in the EP. Where necessary EPs were transferred to another panel.

- 618 At this stage the TEC secretariat also identified EPs where it was necessary to obtain additional input into the assessment. Under the *Guidelines* a panel could obtain additional input into the assessment of an EP when the members of the panel could not provide all the expertise necessary to fully review an EP. There were two sources of additional input:
- Another panel (cross-referral); and
 - A specialist adviser.
- 619 Under the *Guidelines* specialist advisers could be used by peer review panels if:
- subject relevant expertise was not represented on the panel;
 - conflict of interest ruled out the use of panel members with the relevant subject expertise; or
 - members of a panel with the relevant subject area expertise could not reach a consensus on the scoring of components of an EP.
- 620 In practice 87 EPs were referred to specialist advisers.
- 621 Under the *Guidelines* input from another peer review panel could be obtained 'typically' when a significant proportion of a cross-disciplinary RO, but not a majority, fell within the subject areas covered by another panel. In practice 485 EPs were cross-referred in this way. TEOs were advised of transfers but not of cross-referrals. The nature and timing of the advice about transfers has been a concern raised by TEOs with the TEC and a matter requiring improvement.
- 622 Finally the secretariat assigned the EPs for initial scoring to two members of the panel, one of whom was named as the lead member of the pair. In assigning EPs the secretariat had regard for subject appropriateness, workload and conflicts of interest.

Fitness for purpose of the phase

- 623 This process had the purpose of achieving the best and most appropriate alignment of EPs with the peer review panels and the subject areas allocated to them. It was a necessarily complex process because of the great variety of approaches to departmental and subject arrangements in TEOs, the large amount of interdisciplinary research that takes place, and the necessary simplification of this diversity in arrangements that was required in order to establish coherent peer review panel arrangements.

- 624 The process was basically sound and resulted in no fundamental inconsistencies that could not be mitigated by modification. However interdisciplinary and subject boundary questions mean that a fully satisfactory allocation of EPs will never be attained.
- 625 Interdisciplinary research is difficult to assign to established research fields. This difficulty has proven to be an intractable problem for the UK RAE. A process that allocates research funding to institutions on the basis of quality measurement will need to specify categories of research activity where the results of like work can be compared and funding decisions made. But the fields of academic research are increasing in number, complexity and method such that the assignment of all research to a manageable number of categories is either not possible or is clearly irrational. In a funding allocation system based on quality measures, there will always be a need to make uneasy categorisations of some individual researchers or types of research.
- 626 The best approach appears to be to declare the fallibility of the assignment process and to seek to minimise the number of such cases and any distortions in funding that may result. In the UK, consideration is being given to requiring tertiary institutions to name emerging inter-subject and inter-disciplinary fields of inquiry to assist funding allocation processes to take account of the changing research landscape.
- 627 In his *Review of Research Assessment*, Sir Gareth Roberts (Roberts, 2003) considered this difficulty and suggested that overall fairness in the allocation of funding should take precedence over the provision of public information about research in that jurisdiction. His view suggests that in the PBRF a reasonably fair mechanism for allocating funding to TEOs is possible because across a broad range of disciplines and sub disciplines inequities due to categorisation difficulties are likely to produce minimal net distortions. Valid fine-grained comparisons of research performance in disciplines and sub-disciplines are much more difficult to achieve because descriptive categories will always have inbuilt artificial boundaries that do not reflect the true richness and diversity of the research environment. Fair comparisons of individual performance are quite impossible in such circumstances.
- 628 Which academic fields (subjects) are grouped into which panels and which EPs are allocated to those panels was always understood to be an inexact process. Clear boundaries rarely exist and no manageable panel process can represent every sub-discipline, particularly in our small academic community.
- 629 In this report we have accepted the view that the priority for any improvement proposed must be to ensure that the funding allocation is reasonably fair and robust. The perfect assignment of research will never be achieved because of the constant changes that take place in the activities of academic research and the necessary limits to the

Quality Evaluation structures and processes. But fairness and robustness can be improved and demonstrated by continuing to review and refine the peer review panel structure and membership and the subject categories used for allocating EPs and reporting results.

- 630 Accepting an incremental improvement approach to the assessment process may mean that the funding allocation is broadly accepted by the sector as being 'fair enough', but subject comparisons and comparisons over time will always need to be reported with caution and qualification. In any event if it is the case that research is undergoing ever increasing change, seeking comparisons over time may simply diminish in value. Preserving the ability to make comparisons over time from within the funding allocation mechanism might have the perverse effect of encouraging management practices in TEOs that constrain innovation in research subjects and methodology.

Errors and distortions in this phase

- 631 There was evidence of a number of types of distortion in the results that can be explained in part by the decisions made concerning the allocation to panels and subject areas, and the decisions as to whether or not to cross-refer EPs or obtain specialist advice. These include:
- decisions made about where to allocate EPs at the subject boundary where there was ambiguity. An example is the decisions of some staff members to submit their EPs in the subject areas ecology and evolution, whereas they could justifiably have been submitted under the subject agriculture; and
 - the absence of disciplinary or sub-disciplinary knowledge on a panel where EPs were nevertheless not cross-referred, could not be cross-referred or where specialist advice was not obtained. Examples, such as in the case of agriculture, were noted in panel reports. However we also observed examples of sub-disciplines which were discussed as being possibly disadvantaged by the panel, but these discussions were not reflected in the panel report. An example of this was the case of medical statisticians (although other factors, such as their limited role in many large team projects, were also regarded as contributing factors). Another example was that of commercial law, a field that was under-represented on the panel.

Improving the processes in this phase

- 632 The grouping of subjects in panels, and the allocation of EPs according to those groupings, will always be an inexact process. Clear boundaries rarely exist, and no manageable panel process can represent every sub-discipline.

- 633 We believe that the priority for any improvements must be to ensure that the funding allocation is reasonably fair and robust. Our judgment is that this can be achieved by a process of continually reviewing and refining the peer review panel structure and membership and the subject categories used for allocating EPs and reporting results. However, for the reasons outlined above, a stable situation cannot be anticipated.
- 634 While such an approach can lead to a funding allocation that most will be satisfied with, it is less likely that subject comparisons and comparisons over time will ever be able to be reported without caution and qualification.

The assessment of EPs by panel members and the analysis by the TEC

- 635 In this evaluation we observed the entire panel processes but we did not directly observe the initial scoring of EPs by panel pairs. Our inferences about this work were made from the sources of data listed in paragraph 8. We draw on the interview data, observation of panel members reporting on their work in pairs, observations of their work at the meetings of full peer review panels, their responses to the on-line survey and analysis of documents and reports.

Implementation of the phase

- 636 The panel member pairs were first required to score each of the three components of the EP (RO, PE, CRE) separately on a 0-7 point scale. Tie-point descriptors were provided for the scores 2, 4 and 6. A score of 0 was awarded only if no evidence was supplied.
- 637 The scores for each category were then awarded a weighted score in the proportion 70-15-15 between the three components. A maximum score of 7 for each component translated into a maximum available score of 700.
- 638 To assist in reviewing EPs the paired panel members could examine at least one NRO from each EP. The purpose was to assist in making the assessment. They were expected to do so if:
- there were serious doubts about what might be the appropriate score for the EP;
 - a significant proportion of ROs were non-quality assured;
 - additional questions were raised by the examination of a particular NRO; or
 - cross-referral made it seem prudent to examine an NRO.



- 639 *The 2003 Assessment* notes that panel members examined 29% of all NROs. Most panel members we interviewed felt that they needed to examine at least one NRO from each EP in the interests of fairness. An issue in the 2003 Quality Evaluation was the uneven workload between panels and between panel members. The large proportion of EPs examined exacerbated the uneven workload.
- 640 The weighted scores were used to produce an indicative Quality Category with boundaries set at 200 (R/C), 400 (C/B) and 600 (B/A). This translation was undertaken by the TEC.

Fitness for purpose of the phase

The components of EPs and the weighted scores attached to them

- 641 A high percentage (90.0 %) of on-line survey respondents agreed that EPs provide a valid measure of individual research performance⁴. While there was a high level of agreement among the on-line survey respondents that ROs are valid indicators of research quality (92.8 %), lower levels of agreement were noted regarding PE (79.3 %) and CRE (71.4 %).
- 642 Although a fairly high percentage agreed that the combination of ROs, PE and CRE is a valid indicator of research quality (81.0 %), only 63.3 % agreed that the weighting between ROs, PE and CRE 'is about right'. The open-ended comments reveal contradictory views over how the weightings might be adjusted however. Some reported a desire for greater emphasis to be placed on PE while others argued for less.
- 643 It was not a function of this evaluation to determine the answer to this problem. Further information should be gathered in Phase 2 of the Evaluation Strategy in order to address this issue.

Translating scores into Quality Categories

- 644 Although a high percentage of on-line survey respondents agreed that the processes by which Quality Categories and scores were assigned was fair, only 58.35 % agreed that it was actually appropriate to translate scores into Quality Categories (grades).
- 645 Only 40.7 % agreed that these Quality Categories allowed for adequate differentiation of the range of research quality in the EPs submitted. The low level of agreement appears to reflect a general concern that was evident in the open-ended survey responses, as well as in

⁴ The response rate for the survey was 83.3%.

⁵ Many of the variables yielded percentages of 85% or higher, making 58% a relatively low figure.

interviews with panel members and in other evaluation data. One respondent noted:

The difference on the boundary between A&B or B&C grades is almost insignificant yet a line is drawn and a negligible point of difference translates into a whole grade of difference.

646 We indicated in this report that broadly banded Quality Categories are fit for the purpose of generating a funding formula for TEOs, but are not fit for the purpose of being statements about individual performance.

647 In summary:

- the concept of scoring against the three components was generally considered sound;
- the weighting of the three components was not widely supported, and produced some unintended effects; and
- the translation of the scores into Quality Categories caused widespread unease, primarily because of the coarse differentiations that resulted. This appeared to be associated with the widespread understanding of the Quality Categories as being individual grades.

648 It is our view that the scoring process was on the whole sound (with the reservations noted elsewhere in the chapter), but that many eligible staff, and the TEC itself, did not fully appreciate how significant the scores for the PE and CRE components would become at the boundaries between Quality Categories. It was intended that the holistic judgement would be more significant than the scores, but our observations of the panel processes indicate that greater weight was placed on the scores than had been anticipated.

649 The translation of scores into broad bands of Quality Categories serves the process well if it is well understood that the purpose of the Quality Evaluation is only to generate a funding formula for TEOs. But it was not so understood by many during the 2003 assessment. It was viewed by many as a personal performance appraisal, and as such the broad Quality Category bands were seen as providing inadequate differentiation.

The definition of research and its interpretation by the peer review panels

650 In *Investing in Excellence* the PBRF Working Group concluded that the primary focus of the PBRF should be on 'revealing and rewarding researcher excellence and excellent research' (*ibid.*:7). A central aim was to increase the average quality of research (Ministry of Education and Transition Tertiary Education Commission, 2002, p7).

651 The definition of research developed for the PBRF drew heavily on definitions employed by the New Zealand Qualifications Authority, the UK RAE and the OECD (Tertiary Education Commission, 2004a, p22).

652 The PBRF definition states:

For the purposes of the PBRF, research is original investigation undertaken in order to gain knowledge and understanding.

It typically involves enquiry of an experimental or critical nature driven by hypotheses or intellectual positions capable of rigorous assessment.

It is an independent, creative, cumulative and often long-term activity conducted by people with specialist knowledge about the theories, methods and information concerning their field of enquiry. (Note: The term "independent" here should not be construed so as to exclude collaborative work.) Its findings must be open to scrutiny and formal evaluation by others in the field, and this may be achieved through publication or presentation.

In some fields, the results of the investigation may be embodied in the form of an artistic work, design or performance.

Research includes contribution to the intellectual infrastructure of subjects and disciplines (eg dictionaries and scholarly editions). It also includes the experimental development of design or construction solutions, as well as investigation that leads to new or substantially improved materials, devices, products or processes.

653 The intent of the PBRF was to measure:

the quality of the full range of original investigative activity that occurs within the sector, regardless of its type, form, or place of output' so that valid comparisons could be made among subject areas within and among the panels (Ministry of Education and Transition Tertiary Education Commission, 2002, p8).

654 The Working Group report also concluded that:

'excellence' as a researcher in the tertiary system was not just about the production of well-respected articles, books and other forms of research output'. It includes all of a range of activities from the production and creation of leading-edge knowledge; its application and dissemination; as well as supporting current and potential colleagues (Ministry of Education and Transition Tertiary Education Commission, 2002, p7).

655 The observation and interview evidence indicated that interpretations of the tie-point descriptors and the *Guidelines* were sometimes at odds with this definition of research and research quality intended by the PBRF policy. This was exacerbated by other factors, specifically the tight timeframe; high and uneven workload between panels, and between panel members; and the newness of the process. Our findings are that:

- the *Guidelines* offered generic, and some panel-specific criteria;
- although panels were encouraged to develop specific criteria, few did so in an explicit process;
- panels appeared to have developed 'rules of thumb';
- there appeared to be a lack of consistency in the extent of development and application of proxies;
- when combined with the influence of discipline-specific cultural norms, the differences in approach sometimes had adverse effects on within-panel and cross-panel consistency;
- 'calibration' did occur, within and between the panels, in a systematic and rigorous Moderation Panel process ; and
- this moderation process made it possible to discuss some of the proxy indicators of quality more openly. However, there were issues of concern with the moderation process, which we discuss below.

656 In our panel observations, including those in the Moderation Panel meetings, we formed the view that cultural norms associated with particular subjects and institutions were sometimes taken-for-granted in interpreting the descriptors and *Guidelines*, rather than made explicit and subject to more critical review. By cultural norms we include the practices that are typical of particular subjects - these might include epistemological (what counts as warrantable knowledge); methodological (how the knowledge is obtained); the particular stage of a linear model of research ('pure' or 'basic', 'applied'); the form in which the research is made public (journal articles; books; PhDs; original designs; musical compositions; etc), and how the research is valued, or acquires and conveys status (academic peers, industry or community uses of the research and so on). They also include the cultural norms that are attached to research being carried out at a university, college of education or polytechnic.

657 Further work needs to be undertaken on these questions in Phase 2 of the Evaluation Strategy. In our view, a critical question that needs to be addressed is:

'Of the areas that did poorly in the 2003 Quality Evaluation, was this

because the research done in these areas was of low quality, or was it because of the application and interpretation of the descriptors by the peer review panels and Moderation Panel'?

- 658 The results of this evaluation suggest that in some instances where international comparisons and benchmarks were relatively straightforward, the peer review panel process may have favoured traditional disciplines. In these cases the quality of the research 'outlets' and indicators of peer esteem are clearly established, widely known and agreed upon. A related issue for which there are also early indicators is that applied research, especially on New Zealand or Pacific issues and published in the Asia-Pacific region, may have been rated less highly in particular instances.
- 659 The need for robust measures of research quality was highlighted as a key concern both to the TEC directly and to the evaluation team in interviews and written submissions during the period between August and December 2003. The concern reflected the experience of project managers and academic staff in using the *Guidelines* in the internal TEO assessment process (including senior staff who had chaired internal panels).
- 660 The TEC carried out a series of stakeholder visits and discussions with all tertiary education organisations in December 2003. An internal report on those visits was written. It concluded (*inter alia*) that the main difficulty in acceptance of the research definition within the TEOs arose in cases where the meaning of the definition had not been fully understood, or there was a perception that non-traditional research, while counting as research, may not be considered favourably by panels. The report also noted that notwithstanding acknowledgement by TEO staff that non-traditional research was explicitly included in the PBRF research definition, there were still some concerns that this may not be considered equally alongside traditional research.
- 661 The internal TEC report concluded that these were both aspects that could be addressed by better presentation of the research definition and its place in the assessment framework.
- 662 There was also a general view that the descriptors needed to be more inclusive. A number of research output types were also considered to be difficult to classify in the list provided. (The examples given were monographs and working papers.)

Tie-point descriptors

- 663 The question of reliability of the process also concerns whether the tie-point descriptors were applied in a fair, robust and consistent manner. Most of those surveyed on-line agreed that the process by which the peer review panel assigned numerical scores (90.7 %) and interim

grades (90.6 %) was fair. Questions were raised, however, about the consistency of grades across disciplines. There are two dimensions to this issue - consistent application of the tie-point descriptors and other guidelines both:

- within panels across the sub-disciplines; and
- in moderation across panels.

664 The evidence from observations and interviews during the peer review panel assessments in November and December 2003 indicates that, overall, the tie-point descriptors provided in the *Guidelines* and used by panel members to assess EPs were comprehensive. However they contained a number of phrases that proved problematic in their interpretation by members of the panels. The most important of these was the phrase 'world class research' used in the tie-point descriptor applied to Level 6 of the NRO points scale.

Errors and distortions in this phase

Interpreting the Evidence in EPs- an overview of key issues

665 The evaluation evidence suggests that panels (and TEOs in their internal processes) appeared to struggle in interpreting the *Guidelines* relating to the descriptor applying to the score of six on the points scale. This said, in part:

The evidence portfolio would be expected to demonstrate leadership and accomplishment in research exemplified by a platform of world class research that includes highly original work that ranks with the best of its kind.

666 There are indications from the Phase 1 Evaluation Strategy evidence that the assessment of EPs by the 12 peer review panels may sometimes have favoured traditional disciplines where international comparisons and benchmarks are relatively straightforward. Similarly, as noted above, there are indications that in some circumstances, applied research, especially on New Zealand or Pacific issues and published in the Asia-Pacific region, may have been rated less highly. Practice-based research may also have been rated less highly in some circumstances.

667 In making these suggestions, we draw on evaluation evidence from observations of panel processes, and from interview data and material for panel reports (both in their draft and final forms), in the context of commentary prior to and following publication of the results.

668 The evidence provides indications of the nature and significance of the problem. It is important to note that the evidence is drawn on to illustrate the problem, but not to draw inferences about its extent (see Chapter 1). The Phase 2 Evaluation Strategy needs to address this issue with appropriate quantitative measures.

669 Panels we observed raised the question of how value is to be placed under the criteria on work that is New Zealand focused and published in New Zealand or regional journals, and individual panel members raised this matter with us in interview or in the on-line survey. Some panels also raised the question in their draft or final reports. Work that was of great significance at a national or regional level was sometimes seen as being undervalued because it was not published in 'world class' journals or by leading university publishers.

670 A respondent to the panel members' survey provides a reflection on the complexity of the issues concerning the interpretation of the evidence in the EPs:

Panel expertise though broadly based did not cover all the research areas for which EPs were received. In [these] cases generic indicators of quality were relied on rather than specific knowledge of the researcher's field. But whether this inflated scores by giving the benefit of doubt or prevented such EPs from scoring highly is uncertain

671 These issues were recognised in the UK RAE. The *RAE 2008:Initial Decisions* circular (RAE 01/2004) states:

Concerns have been expressed that the [RAE] exercise ... does not deal well with applied and practice-based research in particular...(p4)

The main points from the consultation responses are ... that the assessment process should be designed better to recognise excellence in applied and practice-based research, in new disciplines, and in fields crossing traditional boundaries (United Kingdom Funding Bodies, 2004, p5).

672 The New Zealand Royal Society's Social Sciences Committee PBRF Forum (May 2004) also drew attention to this problem in its *Issues Paper* and in workshop discussions (The Royal Society of New Zealand Social Sciences Committee, 2004, p9).

673 The problem has two components in terms of fairness and consistency:

- the criteria used to specify categories might have been interpreted as giving primacy to the global reach of the research over the quality even where the *Guidelines* were explicit that this was not the case;
- or

- if the status of the journal is used as a proxy for the quality of the research, then it is difficult to identify excellent research that is published in local and regional journals.

674 An example of the issues that this raised is expressed by one of the on-line survey respondents in referring to the work of his/her specific panel. The respondent noted concern about the experienced difference between the *Guidelines* and communications with the TEC, and the experience of actual scoring of evidence portfolios:

In particular, the definition of 'world class' research did not seem consistent. Thus applied research, especially on NZ or Pacific issues published in the Asia- pacific region, was not rated highly. This is not a good way to proceed for NZ. World-class research was defined as that published in the top ranked journals; again something that TEC did not seemingly intend given my many conversations with them on this matter.

675 During observations of panel processes, we also observed confusion between forms of research (along the basic, applied and practice-based continuum) and the quality standard (the tie-point descriptor for 6). For example, a draft peer review panel report, which was written in December 2003, noted the following comment in relation to an applied sub-discipline being considered by this multi-disciplinary panel:

Unless something is published in an international journal (or one known by members of the particular panel to be a 'well-respected journal') it is not able to attain the quality standard of 'world class'.

676 While this comment did not appear in the final panel report, it suggests something of the struggle that multi-disciplinary panels sometimes had in interpreting the *Guidelines* in practice.

677 During panel observations we noted that in some discussions of EPs panel members articulated assumptions about the spread of ROs. In some cases panel members assumed that absence of obvious specialisation as demonstrated by the RO topics was an indicator of shallow scholarship, but others regarded the same pattern as being indicative of a really top class scholarly mind. In one panel this issue was specifically discussed. This led to questions about whether the panel could deduce that an NRO was an 'orphan' or the culmination of a long strand of research work. These questions appeared to remain unresolved in this panel.

678 There were also some discussions about perceived lack of clarity about how to give relative weight to work at different points along the

'paradigm challenging' – 'assimilation into paradigm' – 'work of synthesis' continuum.

- 679 We now turn to indicators of other possible errors and distortions in interpreting the evidence in EPs.

Measuring the quality of practice-based research

- 680 All subject areas where there is a significant element of practice-based research may have suffered in the assessment process in slightly different ways. Panel reports and/or panel members dealing with Architecture, Engineering, Law, Design, Medicine, Public Health, Nursing, Dentistry and Education noted variations of this theme, with the consequence that aggregate measured performance may have been depressed in those fields. Other practice-based subjects such as music, management and marketing may have suffered in a similar way.
- 681 It is reasonable to infer that many PBRF eligible staff who are employed in certain subject areas are expected to be 'practitioners' (e.g. musicians, surgeons, physicians) and thus have less time for 'research'; also, many staff in the medical area are part-time academics, and are primarily employed to teach rather than to conduct research. Similarly, those who combine research with practice and practice-based teaching, tend to devote less time to preparing papers for publication in leading journals. Their research tends to be aimed at improving professional practice.
- 682 Similarly, we were made aware of some significant research that was generated by the need to solve a practical problem in the field, with the solution of the problem rather than publication, being seen as the proof of quality by the researchers.
- 683 Increasingly, academics in TEOs are being enjoined to work collaboratively with industry or the community and the products of their research are sometime constrained in terms of publication.
- 684 These different issues warrant further attention in the Phase 2 Evaluation Strategy.
- 685 In some fields, research is also undertaken 'within the walls of' eligible institutions by practice-based researchers working under contracts that did not meet the staff eligibility criteria of the PBRF. Whether these staff would have scored poorly or highly cannot be inferred from the evaluation evidence.
- 686 A particular aspect of this issue is the measurement of the quality of clinically oriented research.

- 687 The Medicine and Public Health (MED) panel noted in its final report that clinically-based researchers generally scored lower than laboratory-based medical research. This produced expressions of concern that universities would become less willing to put their people into clinically-based research programmes under the PBRF. The panel generally had the view that this reflected the international reality, and that the New Zealand results were similar to those in the UK RAE.
- 688 The panel expressed the view that some scored less well because many 'researchers in the clinical medicine are often working in research in a part-time way because they combine their clinical work with teaching and research, which makes it difficult for them to get top of the range scores.' We suggest that this matter be investigated in the Phase 2 Evaluation Strategy.

Measuring the quality of inter-disciplinary research

- 689 We have addressed issues relating to inter-disciplinary research in early sections of this chapter. The following comment from an on-line survey respondent illustrates further dimensions of this and other issues in interpreting the *Guidelines* in practice:

The notion of quality across disciplines posed difficulties [in the panel process] forcing evaluations to rely on quantity as a default. In general I am concerned that too much emphasis was placed on quantity. Disciplinary specificity was difficult to determine or uphold in disciplines that are under-represented or emerging, that are sub-disciplines of established disciplines, or that are sub-disciplines and inter-disciplinary in nature.

Measuring performance and creative arts

- 690 Further illustrations of the complexity of the issues raised in this section is provided by a member of the Creative and Performing Arts panel in responding to the web-based survey. The member noted that the panel had a particularly diverse range of disciplines to assess. While recognizing the formal *Guidelines*, interpretations were still needed:

We agreed (after much discussion) that original creative work should be treated as research (or equivalent to more orthodox academic research). One issue that kept recurring was how to distinguish between genuine research outputs in this area and good professional practice.

Measuring the quality of indigenous research

- 691 Concerns were expressed by all three wānanga that the overall PBRF design does not take account of Māori epistemology, or the stage of development of wānanga as degree-granting organisations (see TEC internal report on visits to TEOs, January 2004).
- 692 This view was expressed specifically in the form of a paper to the Minister from Te Wānanga O Aotearoa in mid-August 2003, but was a view that was also held by the other two wānanga, and reinforced through the case study data gathered for this evaluation.
- 693 There are specific difficulties in gathering and assessing evidence of research quality relating to indigenous research. These appear to include:
- the difficulty of finding suitable peer-reviewed world class journals in which indigenous research might be published;
 - a significant proportion of indigenous research involves rediscovering lost or almost lost knowledge from community members who hold it, or remember it;
 - the inappropriateness of many conventional research methodologies for gathering and recording knowledge in indigenous cultural settings;
 - the significance of oral traditions;
 - the fact that for many research topics rigorous peer review is to be found in the wider community rather than in the academic community;
 - the need for a significant proportion of indigenous research to be of demonstrable applied utility to the community (in common with other forms of research illustrated in this section of the report);
 - the complex, cultural and political barriers that sometimes induce indigenous researchers to be reluctant to self-promote in 'conventional' academic contexts; and
 - the occasional presence of ethical barriers to the publication of field data (in common with other forms of research).

Improving the processes in this phase

- 694 Most of the potential for improvements in this part of the process have been discussed elsewhere. The potential for more consistency and robustness lie in the tie-point descriptors, their interpretation by panel

members who are better aligned with the EPs to be assessed and in the moderation processes.

The assessment of EPs by peer review panels

Implementation of this phase

- 695 Panels made a holistic judgement to finalise the placement of EPs into Quality Categories using the following information:
- the three scores awarded;
 - the indicative Quality Category assigned by panel members, usually in pairs;
 - notes indicating uncommon factors contained in the EP; and
 - the overall information in the EP.
- 696 Certain rules constrained the final assignment of Quality Categories by peer review panels. These were:
- An RO score of at least 2 was required to award Quality Category C; and
 - An EP did not meet the minimum requirements for Quality Category C if the only NRO was a masters or doctoral thesis.

Fitness for purpose of the phase

- 697 As we discussed in the preceding section, the panel members, together with the TEC, produced an indicative Quality Category based on weighted scores.
- 698 The peer review panel meetings were designed to achieve:
- effective 'calibration' of disciplines within the multi-disciplinary panels; and
 - moderation between the pairs of panel members that initially scored the EPs. This was achieved in part by the deliberations of the panels, but also by having each panel member work with most others in scoring at least one EP.
- 699 Panel chairs and members of the TEC secretariat supported this process by commenting on deliberations, based on interpretations and judgements that had been reached in Moderation Panel meetings (see below).

Errors and distortions in this phase

Summary

- 700 As we have argued elsewhere, any assessment of research quality is a partly subjective exercise since it is mediated by cultural norms, values and practices, and is subject to human error. Accordingly, the evaluation paid special attention to whether there was evidence of systemic bias in the assessment and moderation processes, both within and between multi-disciplinary peer review panels.
- 701 Overall, we found that the peer review panels worked very hard to treat EPs fairly and consistently. They worked within a very sophisticated process designed to help them achieve this. In observing many hours of work in five panels as well as the Moderation Panel we saw an extraordinarily high level of commitment to apply the processes professionally, fairly and rigorously.
- 702 In seeking to achieve this, the panels engaged in frequent sessions of self-reflection about their own processes, their own perceptions and their own sense making.
- 703 Nevertheless we did observe examples of unconscious biases that were not always corrected by the checks and balances in the process.
- 704 We found that the extent to which unfairness and inconsistency occurred was due to one or more of the following factors:
- ambiguities in the tie-point descriptors;
 - the use of proxy evidence of quality because of the impossibility of sighting all the NROs themselves and the ways in which such evidence was understood;
 - poorly completed EPs that provided inadequate or ambiguous evidence;
 - the unconscious use of personal knowledge, which did not amount to conflict of interest, about some of the academics whose EPs were assessed;
 - panel workload overall, and variability in workload within and among panels;
 - divergent working practices between panels;
 - in some panels there was inadequate knowledge about some of the subject areas being discussed;

- a predisposition to retreat to personal mental models of what constitutes research and quality in the face of ambiguous or inadequate evidence. These mental models tended to reflect implicit cultural norms relating to particular disciplines, academic units or institutional types; and
- the influence of a particular panel member.

Impact of the scoring system on new and emerging researchers

- 705 The most common remark made in response to the first open-ended question was directed at the Quality Categories (grades) and their relevance for new researchers. As one respondent remarked: 'New researchers in the first few years of their careers are certainly disadvantaged in terms of PE and CRE compared with more established researchers'.
- 706 Such observations characteristically led to some rather harsh criticism of the 'R' grade in particular with many advocating, for example, an 'N' rating for new researchers.
- 707 Panel members' views on this issue were consistent with the universal view that the generic descriptors for the category R were wrong since they defined postgraduate researchers, and recently-appointed university academics as 'Research Inactive'.
- 708 The use of 'proxies' for quality is unavoidable in an assessment exercise such as the Quality Evaluation. In the panel processes we observed we saw a working understanding of how proxies for measures of quality were to be interpreted emerge. This was particularly the case for articles published in journals.
- 709 Panels we observed were cautious about providing quality assurance outside of their field, especially NROs that may have rated very highly. We observed the emergence of an informal hierarchy of rated journals. An NRO published in a highly regarded, high impact journal was not simply taken as quality assured, it appeared to be given extra weight in the overall assessment process if the other NROs were weaker.
- 710 Our interpretation is that this may have introduced an inflationary drift in the determination of scores for the RO component that favoured one particular view of what counts as quality and how much it counts. An alternative view is that this recognition is appropriate when a first-rate piece is published in a top-quality journal.

Holistic judgements

- 711 In the final step of placing EPs into Quality Categories the *Guidelines* instructed the peer review panels to have regard to the scoring of an EP, the information contained within it, and any special factors relating to

the EP. In practice holistic judgements by panels resulted in changes to the categories indicated by the scores in the range of 2%-5% across all panels. Holistic judgements were necessarily more subjective than was the scoring of EPs.

- 712 The *Guidelines* stressed to peer review panel members that the assessment was of research quality, not quantity. However the *Guidelines* also indicated that EPs needed to demonstrate that there was an adequate 'platform of research' to justify Quality Categories. We observed that the peer review panels we visited gave more weight to the quantity of research than the *Guidelines* intended, and paid special attention to the number of research outputs (ROs) other than NROs listed in an EP.
- 713 The greater than intended weight given to the quantity of NROs and ROs means, in our judgment, that the application of the special circumstances provision had a different effect than intended. Some panel members felt that the use of special circumstances might have rewarded TEOs that were administratively top heavy. Without closer analysis of the results it is not possible to reach a firm judgement whether or not this is so.
- 714 One example was acceptance of the view that holding an 'Oxbridge' Ph.D. was, *per se*, evidence of higher quality than a Ph.D. from another university. It followed that the RO component score was higher, and that the EP received a higher Quality Category.
- 715 Although the integrated *Guidelines* for panel members specifically stated that: "Research outputs that deal with topics or themes of primarily local, regional or national focus can be of world standard", in practice peer review panel members often found it difficult to judge whether this was so for any particular NRO unless it was one that they personally sighted and had personal subject expertise about. Proxy evidence used by panels did not always allow the panel to make an accurate judgement. Some panels and many panel members specifically identified this as a problem.
- 716 The most commonly used proxy indicator of quality was the international standing of journals. While publication in a leading journal could be taken as evidence of world class research we observed that panel members were often confronted with the problem that publication in a lesser journal, or a New Zealand journal, could not necessarily be taken to indicate the lack of quality. In the absence of clear positive evidence of quality panel members often scored an NRO as though the journal in which it was published represented its quality. This problem and the solution adopted was observed by us in every panel we visited, and described in the report of a panel we did not visit.

717 The tie-point descriptors pointed to other possible indicators of quality than journal publication, but in practice peer review panels were not always able to deduce the quality of the NRO from the evidence given. These problems in the use of proxies may explain the relatively poorer showing of practice based research, New Zealand research, applied research, and performance and artistic research. However, further study is required to test this proposition.

718 We also note other issues that arose in the process of interpreting the *Guidelines* in practice:

- the use of the term 'invited paper' to cover a very wide quality range;
- the non specified use of the term 'successful supervision';
- the nature of books for which chapters were cited; and
- where performances occurred.

Conflicts of interest

719 Rigorous procedures were used to eliminate conflicts of interest in the peer review panel process. What constituted a conflict of interest was clearly defined in the *Guidelines*. In the panels whose work we observed these were meticulously adhered to under the guidance of panel chairs and the secretariat. This was absolutely necessary in New Zealand's small academic community. Panel members surveyed, and those interviewed, were almost entirely satisfied that they had behaved properly and applied the *Guidelines* effectively.

720 However it is also the case that in the New Zealand academic community it is unusual to have no personal knowledge at all of other academics working in the same subject field. We observed many examples of the unconscious and unintended use of personal knowledge, not constituting a conflict of interest as defined in the *Guidelines*, that introduced evidence not included in the EP. On occasions that personal knowledge was revealed when a panel member expressed their frustration at person who had submitted an EP and had not done him or herself justice.

721 Panel chairs were alert to the use of personal knowledge, but we also observed occasions when it seemed to us that personal knowledge not contained in an EP was introduced as evidence, but was not ruled out by the panel chair. Some panel members, and at least one panel chair, were concerned that the use of personal knowledge was sometimes used to advantage (or disadvantage) an EP that was discussed by the panel, whereas many researchers whose EPs were not discussed by the whole panel were not exposed to the same processes.

- 722 The following quotations from on-line survey respondents are good examples of how distortions may occur:

Individual panel members influenced evaluations within their discipline unduly.

It is difficult to assess a colleague in the same Faculty [and]...it is almost impossible for the assessor to demonstrate that there is no bias.

[Evaluating other panel members' EPs] was very awkward and it was clear that people felt reluctant to argue against upgrading others in the group in case the same arguments were used against them.

- 723 Some, but not all, international panel members that we interviewed reported that they monitored the use of personal knowledge by panel members, and alerted their panels when they observed the apparent use of it. However some Australian panel members told us that they felt they the New Zealand and Australian academic communities within their subject area was so intertwined that they experienced almost as many cases where they had personal knowledge as did their New Zealand colleagues.
- 724 The *Guidelines* prepared by TEC to assist panels were clearly important. We saw many instances in which a discussion that seemed to be going off-track was rescued by a panel member referring back to the *Guidelines*. The performance of the panel chairs was critical here. The information given to panel members in the *Guidelines* was the basic source used in making judgments.
- 725 In our observations panel members often forgot the *Guidelines* which was not surprising given the complexity of the task. Good and early interventions by panel chairs or members of the secretariat who had a sound grasp of the *Guidelines* (and such was usually, but not always, the case amongst the panels we observed) was a critical factor in maintaining robustness and fairness, as well as in saving panel time. This usually, but not always, occurred.
- 726 Panels often discussed how they were to interpret the *Guidelines*. This produced varied outcomes. In one panel the decision was to 'err on the side of generosity.', but in another there was an injunction to be 'rigorous and critical'. It is not clear whether this produced cross-panel inconsistency, but there is a prima facie case for using the 2003-4 experience as a basis for making the *Guidelines* more specific.

The provision for special circumstances

- 727 The special circumstances provision provided both design and interpretation difficulties. The *Guidelines* allowed special circumstances to be used to justify fewer than four NROs being included in the EP. The quality standard remained the same. It is our observation that the distinction between quality and quantity was not always an easy one for panels to make.
- 728 The Moderation Panel spent considerable time discussing the application of the special circumstances provision in the EPs and its impact on the work of the panels. Interview respondents generally agreed that this did help to clarify the special circumstance provision, but that its use remained ambiguous. In particular, observation data suggests there were some EPs where a staff member had a substantial administrative workload, but was still very productive as a researcher. Yet allowances were made for colleagues with similar workloads who were not so productive, although inferences cannot be drawn about the impact of this upon the final result.
- 729 We observed that the application of 'special circumstances' during holistic assessment sometimes produced puzzling judgements. This comment by a panel member in explaining the allocation of a score above the B boundary for an EP that had only 2 NROs characterises the problem: 'I took account of the fact that she was on maternity leave and part time for part of the period'. After some discussion the panel accepted the category despite agreeing that the quality of the two NROs alone would not have led them to assign that Quality Category without the application of the special circumstances provision.
- 730 In one case we observed a final decision to raise an EP from a 'C' to an 'A' was made explicitly on the grounds that not to do so would be unfair to the individual. This was based on the person's involvement in community development within his field, rather than on the quality or quantity of research presented as evidence.
- 731 Holding a leadership role in the TEO was sometimes recorded as special circumstances, and sometimes as evidence under PE or CRE.
- 732 We stress that it is inappropriate to draw inferences about either the extent to which these instances occurred during the 2003 Quality Evaluation, or the extent to which they had an impact upon the result.
- 733 Moreover, while these considerations are important with respect to individual fairness, it is also directly relevant to the PBRF policy, which is to allocate funds to institutions on the basis of their overall research productivity and quality. If allowances are made for some individuals because special circumstances impaired their research productivity, then

the consequence may be a distortion in the aggregate results of the employing TEO.

Improving the processes in this phase

734 The evidence in this evaluation leads us to conclude that the foregoing issues are possible problem areas that need investigation and appropriate remedial action before the 2006 Quality Evaluation. However, the methodology does not allow inferences to be drawn about the extent of the problems. These are matters for the Phase Two evaluation design.

735 Our preliminary view is that the panels we observed were negotiating and designing the operational details of the EP assessment process, including measures and proxies for quality and how those would be applied. This is to be expected. We recommend that the TEC develop a strategy to engage with this ongoing negotiation in order to continue the momentum for changed understandings and practice around the assessment of the quality of an EP.

The moderation process

The implementation of the phase

736 The purpose of the moderation phase was to ensure consistency across the peer review panels. Moderation of scoring within panels was built into the panel processes as described above

737 The Moderation Panel consisted of the chairs of the peer review panels and an independent chair.

738 The moderation process had four stages:

- an initial Moderation Panel meeting reviewed the preliminary numerical scores given by panel members. It identified issues that might impact on consistency of standards across panels;
- after the Quality Categories had been assigned by panels, the Moderation Panel met again, reviewed the results and identified panels in which there appeared to be consistency issues;
- panels identified in stage two considered the issues raised by the Moderation Panel; and
- the moderation Panel reported to the TEC Board.

739 In practice, two panels were identified in the third stage. One panel met again and modified its Quality Categories.

Fitness for purpose of the phase

- 740 This phase worked as intended. The Moderation Panel identified two panels where there was apparent inconsistency, but after further investigation eliminated one of these. The other panel met again, reconsidered its Quality Categories and made some adjustments.
- 741 Technically, therefore, the moderation process operated successfully. As indicate below there was some disagreement amongst panel members as to whether the outcomes were what was required.

Errors and distortions in this phase

- 742 The survey of panel members shows that while they had a high degree of confidence that their judgements were consistent within disciplines, they were much more divided about whether there was consistency across disciplines. Questions associated with this were discussed in depth at the Moderation Panel meetings, but ordinary panel members were not party to these discussions.
- 743 In light of such high levels of agreement, it is important to note that only 59.9 % agreed that the moderation process ensured that the allocation of final grades was consistent across disciplines. This relative lack of confidence in the process is reflected in the answers provided in Section 3, Q 3.2, 3.3, 3.4 and 3.5 of the survey (see Appendix 2) While 85.7 % agreed that EPs allow valid comparisons of individual research performance to be made within the same discipline, only 44.3 % agreed valid comparisons could be made across different disciplines. A similar pattern emerged during analysis of levels of agreement with the statements 'Processes associated with the quality evaluations of EPs were consistent within the same discipline' (85.7 %) and 'Processes associated with the quality evaluations of EPs were consistent across different disciplines' (33.1 %).
- 744 A panel member in the on-line, web-based survey, summed up some of the complexity of carrying out the exercise in this response:
- Measuring quality is of course a very difficult task. The [peer review panel] process has done admirably in this but effectively I think panels have had to use proxy measures for quality. These proxy measures may be discipline or panel specific, so I guess the inter-panel moderation process is introducing a lot of subjectivity in the process.
- 745 A contrary view, expressed to us during the evaluation, was that the moderation process removed some of the subjectivity.
- 746 We also note the comment made by the chair of the Moderation Panel in his presentation to the Royal Society PBRF Forum ((Callaghan, 2004, p2):

Panel members cannot help but bring to their deliberation a sense of the value of their respective discipline areas. It is therefore vitally important that our processes ensure that any sense of competitiveness between discipline areas should not play a significant role.

Improving the processes in this phase

747 The chair of the Moderation Panel was able to be present for significant periods of time at 4 out of the 12 panel meetings (Callaghan, 2004, p2).

748 The chair of the Moderation Panel concludes that 'it may have been helpful to have had a greater scrutiny of the individual panel assessment processes by members of the Moderation Panel' (Callaghan, 2004, p2). He recommends that the TEC adopt the practice that either the chair, or other members, could be present at the holistic assessment stage for each panel. We endorse that recommendation.

The recommendation of Quality Categories to the TEC Board

749 The board received the recommendations as to Quality Categories and accepted them. This was a formal process and there is nothing for the evaluation to comment on other than to say that the formal acceptance of the recommendations indicated the TEC Board's confidence in the outcomes of the 2003 Quality Evaluation.

Conclusion

750 We examined evidence in this chapter related to the following key questions when evaluating the panel assessment process:

- Was the definition of research wrongly or inconsistently applied in some subject areas?
- Did any of the panels misinterpret the definition of research?
- Are the current descriptors of research quality too culturally-bound and biased against certain kinds of research?
- Did any of the panels employ any criteria for assessing research quality that were clearly inconsistent with the *Guidelines* (the descriptors and other measures of quality)?
- Did any of the panels ignore or misapply the *Guidelines*?
- Did any of the multi-disciplinary panels apply the descriptors and other measures of quality in an obviously inconsistent manner,



particularly with respect to within-panel and cross-panel comparisons?

- 751 These questions have become significant since the results and analysis of the Quality Evaluation were published by TEC in late April 2004, and the early impacts of those results on TEOs and individual staff have become to emerge.
- 752 This evaluation provides early indicators of issues and areas of concern that warrant further discussion and investigation in Phase 2 of the Evaluation Strategy.

References

- Boston, J. (2002). *Designing a Performance-Based Research Fund for New Zealand: Report for the Transition Tertiary Education Commission*. Wellington, New Zealand: Victoria University of Wellington.
- Boston, J. (2004). *The Future of the Performance-Based Research Fund: Issues and Options*. Wellington.
- Callaghan, P. (2004, 21 May). *Moderating the PBRF Quality Evaluation: Some Issues and come Concerns*. Paper presented at the Royal Society of New Zealand PBRF Forum, Wellington.
- Engestrom, Y. (2000). Activity Theory as a Framework for Analyzing and Redesigning Work. *Ergonomics*, 43(7), 960-974.
- Engestrom, Y., Virkkunen, J., Helle, M., Pihlaja, J., & Poikela, R. (1996). The Change Laboratory as a Tool for Transforming Work. *Lifelong Learning in Europe*, 2, 10-17.
- Guba E & Lincoln Y. (1989). *Fourth Generation Evaluation*: Sage Publications.
- Hazelkorn, E. (2002). Challenges of Growing Research at New and Emerging HEIs. In W. G (Ed.), *Enterprise in Universities: Evidence and Evaluation* (pp. 18). London: SRHE/Open University.
- Higher Education Funding Council for England. (2000). *Review of Research : Consultation Report 00/37*. Bristol.
- Liebrich, J. (1993). *Straight to the Point: Angles on Giving up Crime*. Dunedin: University of Otago Press in Association with the Department of Justice.
- Ministry of Education and Transition Tertiary Education Commission. (2002). *Investing in Excellence: The Report of the Performance-Based Research Fund Working Group*. Wellington.
- Murnane, R., & Cohen, D. (1986). Merit Pay and the Evaluation Problem: Why Most Merit Pay Plans Fail and a Few Survive. *Harvard Educational Review* (February 1986).
- National Library of New Zealand. (2003). *Performance-Based Research Fund Audit: Verification Final Report*. Wellington.
- Roberts, G., (Sir). (2003). *Review of research assessment : Report by Sir Gareth Roberts to the UK funding bodies*: Oxford.
- Rogers, J. (2002). *Research Assessment Exercise 2001 : Manager's Report - May 2002*. Bristol.
- Taylor, N. C., Bryan, C. H., & Goodrich, C. G. (1990). *Social Assessment: theory, process & techniques* (Vol. 7). Christchurch: Centre for Resource Management Canterbury.
- Tertiary Education Advisory Commission. (2000). *Shaping a Shared Vision*. Wellington.
- Tertiary Education Advisory Commission. (2001a). *Shaping the System*. Wellington.
- Tertiary Education Advisory Commission. (2001b). *Shaping the Strategy*. Wellington.
- Tertiary Education Advisory Commission. (2001c). *Shaping the Funding Framework*. Wellington.
- Tertiary Education Advisory Commission. (2003). *PBRF 2003 Research Assessment: Project Manager's Report*. Wellington.
- Tertiary Education Advisory Commission and the Ministry of Education. (2003). *Evaluation Strategy Proposed by the Ministry of Education and the Tertiary Education Commission*. Wellington.
- Tertiary Education Commission. (2003). *Performance-Based Research Fund: A Guide for 2003*. Wellington.
- Tertiary Education Commission. (2004a). *Performance-Based Research Fund Evaluating Research Excellence: the 2003 assessment*. Wellington.



- Tertiary Education Commission. (2004b). *Performance-Based Research Fund Panels Nomination, Selection and Appointment Processes Debriefing'*. Wellington: Tertiary Education Commission.
- The Royal Society of New Zealand Social Sciences Committee. (2004). *PBRF Forum Evaluating the Assessment Framework*. Wellington: The Royal Society of New Zealand.
- United Kingdom Funding Bodies. (2004). *RAE 2008: Initial decisions by the UK funding bodies* (No. RAE 01/2004).



Appendix 1: Phase 1 Methodology

Summary

- 753 In September 2003, the TEC contracted the Centre for Work, Education and Business Ltd (WEB Research) to undertake a part of Phase 1 of the Evaluation Strategy of the PBRF.
- 754 Phase 1 of the Evaluation Strategy focused on the Quality Evaluation framework developed by the TEC to measure research quality, and whether the framework had been implemented by the TEC and TEOs in a robust, consistent and fair manner. In particular, the evaluation focused on design issues; and the preparation, processing and assessment of Evidence Portfolios (EPs) by the relevant parties (i.e., individuals, TEOs, the TEC, and the peer review panels). This included both the internal assessment of EPs by TEOs and the external assessment by the peer review panels, including the Moderation Panel.
- 755 This evaluation comprised an overall case study in two stages:
- **Stage One: Scoping**

Clarifying the objectives, research questions, evaluation constraints and design, based on interviews and review of documents.
 - **Stage Two: In-depth Study**

The second stage included a further 6 components. These are outlined in Table 1 below. Table 1 also shows the 4 units of analysis used in the research.

Table 1: Overall Case Study Components: Units Of Analysis, Methods, Sources, Sites Of Data-Gathering

CASE STUDY COMPONENTS UNIT OF ANALYSIS	1. Scoping ¹ , Design, Sector Engagement	2. Analysis of Direct and Indirect Costs		3. Case study of Internal Assessments - 3 TEOs (university, college of education and wānanga)				4. Case study of Panel Assessments (MOD Panel; EDU, HAL, MKD, MEDPH)			5. Assessments of RDC and ERI	6. TEC Results, Responses and Impacts			7. Synthesis and Report			
		<i>Analysis of direct costs</i>	<i>Case study-based data</i>	<i>Interviews</i>	<i>Document analysis</i>	<i>Workshop</i>	<i>PBRF Proj Mgrs Group teleconference</i>	<i>Observations²</i>	<i>www survey of Panel Chairs</i>	<i>Analysis of results by the TEC & MoE</i>		<i>Sector Pre³</i>	<i>Submissions Post⁴</i>	<i>Analysis of results of Quality Evaluation</i>	<i>Post-result phone survey of 20 Research Mgrs</i>	<i>Integrate Data</i>	<i>Workshop findings</i>	<i>Recommendations</i>
1. Design criteria, implementation of PBRF Framework		✓	✓	✓	✓		✓	✓				✓						
2. Internal TEO assessments			✓	✓	✓	✓	✓			✓								
3. National Panel assessments								✓	✓	✓								
4. Results and impacts												✓	✓	✓				

CODE: # A tick in a particular column indicates the source, site and method of data-gathering for relevant evidence.
 + Unless otherwise shown, the evaluation tasks are carried out by the independent evaluators.



NOTES:

- i. See Appendix 1. Includes pre-September emails and written submissions to the TEC, newspaper articles, and the scoping interviews.
 - ii. Panel processes: Observations of: a) the TEC's design of processes; b) the TEC's management of; c) Panel Chairs and members' processes for assessments.
 - iii. Submissions to the TEC prior to independent evaluation.
 - iv. Submissions to the TEC following the public release of results which will be administered by the TEC and may be analysed by WEB Research in the early stage of the Phase 2 Evaluation Strategy.
- 756 Because of the delay in publication of the results, some research tasks had to be postponed until Phase 2 of the Evaluation Strategy.

Evaluation Questions

- 757 Phase 1 of the Evaluation Strategy focused on the following questions:
- Whether the design of the PBRF framework appears to be appropriate?
 - Whether the assessment of EPs was conducted in a robust, consistent and fair manner?
 - How the TEOs and individuals responded to the quality evaluation process and why?
 - What were the direct and indirect financial costs of the implementation of the PBRF?
 - The results of the 2003 Quality Evaluation.

WEB Research evaluation design

- 758 This evaluation design involved an overall case study of the implementation and early impacts of the PBRF and the 2003 Quality Evaluation process. This evaluation had seven components. These are shown in Table 1 below.
- 759 An initial scoping component was needed because of the policy environment, multiplicity of stakeholder interests and complexity of data sources. The intent was to clarify the objectives, research questions, evaluation constraints and design, based on interviews and review of documents. WEB Research carried out the scoping component between September and November 2003.



- 760 An in-depth study with six further components was conducted between December 2003 and June 2004. It involved:
- an analysis of direct and indirect costs;
 - case study of internal assessments in 3 TEOs;
 - case study of panel assessments involving detailed observations of four panel, semi-structured interviews during the panel process and a web-based, on-line survey of panel chairs and members;
 - assessments of RDC and ERI by the MOE;
 - the TEC results and impacts; and
 - synthesis of the case study data and a report
- 761 Those who are unfamiliar with case study methodology are likely to be concerned at the apparently small number of case studies proposed for components 3 and 4 (Table 1). When these case studies are combined with the other overall case study components, including the submissions from TEO's and individual academics, they are expected to produce a solid basis for Phase 1 recommendations and/or for further in-depth study in Phase 2 of the Evaluation Strategy.
- 762 Case study research involves triangulating multiple data sources, methods and points in time. Another way of describing this is to talk about analysing 'layered sources of data' (Liebrich, 1993) to reveal patterns and causal connections. The analysis is carried out using analytic-induction techniques. Analytic-induction can be defined by contrasting it with survey research that relies on statistical analysis of quantitative data. It is not the numbers that make data valid under the case study process, but the logical integration of data from different sources and different methods of analysis into a single, consistent interpretation. The integration of both qualitative and quantitative data to determine consistency (and to account for inconsistencies) adds power to the analysis (Taylor, Bryan, & Goodrich, 1990).
- 763 Each of the seven components listed in Table 1 are described in more detail below.

Scoping

- 764 Scoping was carried out in consultation with the TEC Evaluation Advisory Group and involved preliminary interviews with selected TEC, MoE and TEO staff, and sector organisations including the AUS. Further scoping interviews were undertaken with wānanga staff, the Pacific Advisory Group, the Association of Polytechnics of New Zealand, the NZVCC and professional associations.



Analysis of direct and indirect Costs

- 765 *Direct costs.* An analysis of quantitative data supplied by the MOE, on the administrative costs to the MOE and TEC of the implementation of the PBRF and the conduct of the 2003 Quality Evaluation was carried out. The data was interpreted in the context of the qualitative data gathered in the other components of the case study.
- 766 *Indirect or compliance costs.* A sample of TEOs were approached and invited to provide estimates of the cost to them of their compliance with the implementation of the PBRF and the 2003 Quality Evaluation.

Case studies of EP preparation and assessment at three TEO's

- 767 Interviews and exploratory group meetings were held at the University of Otago, the Christchurch College of Education and Te Wānanga o Aotearoa covering a spread of disciplines and panels. This process involved PBRF project and research managers, chairs and members of selected internal panels, a group of academics and the local representatives of AUS, ASTE and other staff associations. The selection of TEOs was based on consultation with those interviewed in the scoping exercise, within the context of Cabinet requirements.
- 768 Data-gathering focussed on the five evaluation questions in the context of how the internal assessment processes operated at the individual TEO level, and included:
- the nature, structure and functioning of the internal panels;
 - the nature of the information provided to staff; and
 - the way TEOs dealt with those judged to be in the 'R' category.
- 769 A key issue explored here and in the other set of case studies was whether the self-evaluation process could be relied upon without the use of external panels; conversely, whether the self-evaluation process should be abandoned except in particular circumstances, with reliance being placed primarily on the external panels.

Case studies of assessment processes by peer review panels and the Moderation Panel

- 770 Three members of the WEB Research evaluation team carried out observations of four peer review panel meetings and interviewed a small selection of panel chairs and members. Two WEB researchers attended the initial Moderation Panel meeting on 15 November and three attended the second meeting on 15 December. A WEB researcher was

also present at each of the two additional special meetings of the Moderation Panel held in December 2003 and January 2004.

- 771 Four of the twelve peer review panels were selected for in-depth examination on topics relating to clinical education (MEDPH); non-university participation (EDU); Maori research and researchers (MKD); and a broad and diverse range of non-‘science’ disciplinary areas and professional activity (HAL).
- 772 Each of these topics was either a Cabinet requirement, or had emerged in the scoping as a key issue for further investigation. Further data-gathering was carried out through informal semi-structured interviews, email communication, and through the web-based survey (see 2.1.4(i) below). Interviews relating to the panel process were also conducted with sector organisations including the Pacific Advisory Group, the NZVCC, and the national offices of the AUS, ASTE, and other staff associations.

Table 2: Interview Summary

Panel	No. Days	No. Researchers	No. Interviews
MOD	2	2	2
EDU	3.5	2	2
HAL	2	1	6
MKD	1	1	3
MEDPH	2.5	1	6

- 773 Data-gathering at the TEC and in the case studies of TEOs explored:
- Chair and panel members’ selection including criteria, process, suitability, balance of subject areas/expertise, gender balance, the adequacy relating to the number of 'independent' overseas members, cost, number of resignations (and reasons), sector concerns.
 - Design of the assessment framework, including assessment criteria, guidelines, and international comparability.
 - The assessment and moderation process and issues
 - The reporting framework and issues.
- 774 The role and concerns of the polytechnics were addressed in this component, including their request for a separate fund and the policy work that has been undertaken by central agencies as a result. Follow-up in-depth interviews were conducted to explore these issues.



Web-based on-line survey

- 775 With agreement from panel chairs and members, and in consultation with the TEC, WEB Research designed an internet-based on-line questionnaire to investigate panellists' views on the PBRF and the 2003 Quality Evaluation. The questionnaire design was based on qualitative data gathered at an earlier stage of the research. We decided to include a range of both close-ended and open-ended questions in the on-line survey. The close-ended questions allowed for a wide range of data to be gathered quickly and facilitated data analysis. Open-ended questions were also included in order to provide an opportunity for respondents to elaborate on any issues that interested or concerned them.
- 776 On the 2nd March 2004, an initial email letter was sent to the respondents advising them of the up-coming on-line survey. This was followed by the first email invitation to participate in the survey and this letter also contained details of the WEB Research team, a brief description of the length and types of questions in the questionnaire and a link to the website where the on-line survey could be completed. A second email reminder was sent on the 14th March and a final reminder was sent to those who had not yet replied on the 20th March. A small number of people encountered difficulties completing the survey on-line and these were sent hard copies which were returned to WEB Research by post.
- 777 A response rate of 83.3% was achieved for the survey.

Assessments of RDC and ERI

- 778 Initial analysis was completed by the MOE. Further analysis is planned for Phase 2 of the Evaluation Strategy.

Analysis of submissions from TEO's and academics

- 779 Given the delay in the release of the results, submissions to the TEC will be invited in Phase 2 of the Evaluation Strategy. The scope, method and timing is expected to be finalised by the TEC later in 2004.

Background to the Evaluation Approach for Phase 1

- 780 The overall evaluation was developmental, using a 'fourth generation evaluation' approach (Guba E & Lincoln Y, 1989).
- 781 The approach also draws on key principles of the Finnish methodology of developmental work research (Engstrom et al., 1996) which are appropriate for contexts, such as that of the PBRF, with the following features (*ibid*):

- a complex, rapidly-changing policy and operational environment;
- a situation where there are multiple voices and actors, multiple agencies and stakeholders;
- an environment that crosses multiple boundaries;
- a situation in which the knowledge (judgements, information and data) that is relevant to the evaluation objectives is held by those who are affected by the policies (in this case, the tertiary institutions, academic researchers, those who represent their respective interests, and the wider New Zealand community); and those who make and implement the policies (in this case, ministers, central policy agencies, policy and operational staff); and
- similarly, a situation in which multiple sources of data, frameworks and literatures are relevant to the evaluation. For example, the evaluation objectives require analysis of data at the national and institutional levels (in this case, held by the TEC, the MoE, and the TEOs), as well as qualitative data obtained at the institutional and individual levels.

782 Analysis, report-writing and recommendations are derived from triangulating the quantitative and qualitative data from multiple sources, methods and points in time to establish robust findings. For example, results of the quantitative analysis of direct costs would be available for triangulation with primary qualitative data on indirect costs obtained in site visits to TEOs.



Appendix 2: On-line survey of Peer Review Panel Members and Panel Chairs

Research brief

783 The On-line Survey was part of the larger case study evaluation of the PBRF and the 2003 Quality Evaluation undertaken by WEB Research⁶. This evaluation focused on the Quality Evaluation framework developed by the TEC to measure research quality, and whether the framework had been implemented by the TEC and TEOs in a robust, consistent and fair manner. The wider study, of which this on-line survey was part, addressed the design of the PBRF framework and the preparation, processing and assessment of Evidence Portfolios (EPs) by the relevant parties. The overall case study emphasised that an understanding of the implications is crucial for ongoing policy development.

784 The on-line survey focused on generating evidence and analysis that related to:

- whether the design of the PBRF framework appeared to be appropriate; whether the indicators of research quality were valid and reliable; and whether the Quality Categories were appropriate discriminators. Of particular note here were early indicators of the success or otherwise of the PBRF evaluation method; issues around what constitutes research and what gets measured within the PBRF; the choice of, and weightings for, the three quality indicators (the Quality Evaluation, RDCs and ERI) and the three components in the Quality Evaluation (RO, PE and CRE); the funding weights; the unit of assessment for the Quality Evaluation (individual versus group); the Quality Categories (A, B, C, R); and the merits of having a dual system of internal and external assessment; and
- whether the assessment of EPs was conducted in a robust, consistent and fair manner; if not, were the problems a consequence of the design of the PBRF, or its implementation?

785 In February 2004, WEB Research and associates undertook the design of the on-line survey, the survey and the analysis and report of the results.

Survey Population

786 It is important to note that the survey population comprised only members of the peer review panels and the results should not be taken

⁶ Much of the information presented in the Research Brief is taken from the WEB Research PBRF Phase 1 Evaluation Plan, 12th December, 2003, p.11.



to represent those of academia more generally. However, as the participants represent a high level of expertise their cautions and critiques are due careful consideration.

Research Method

787 WEB Research was granted access to the email addresses of all panel members so an internet-based, on-line questionnaire was designed to investigate panellists' views on the PBRF and the Quality Evaluation Framework. While the questionnaire design was informed by qualitative data gathered at an earlier stage of the research, it was decided to include a range of both closed- and open-ended questions in the on-line survey. The close-ended questions allowed for a wide range of data to be gathered quickly and facilitated data analysis. Open-ended questions were also included in order to allow for respondent expression and to provide an opportunity for respondents to elaborate on any issues that interested or concerned them.

The Survey

788 On the 2nd March 2004, an initial email letter was sent to the respondents advising them of the up-coming on-line survey. This was followed by the first email invitation to participate in the survey and this letter also contained details of the WEB Research team, a brief description of the length and types of questions in the questionnaire and a link to the website where the on-line survey could be completed. A second email reminder was sent on the 14th March and a final reminder was sent to those who had not yet replied on the 20th March. Two people encountered difficulties completing the survey on-line and these were sent hard copies of the on-line survey which were returned to WEB Research by post. All data were collected and analysed before the results of the first Quality Evaluation were made public.

Presentation of Research Results

789 The results of the PBRF Evaluation On-line Survey are divided into two parts. Part A contains the results of the close-ended questions and Part B is devoted to the presentation of the themes identified during analysis of the responses to the two open-ended questions. In some cases, detailed statistical results have been withheld in order to ensure that the anonymity of the respondents is maintained. In the tables, 'missing' refers to those respondents who chose not to answer. Abbreviations have been used when referring to the different panels:

- BS Biological Sciences
- BEC Business and Economics



- CPA Creative and Performing Arts
- EDU Education
- ETA Engineering, Technology and Architecture
- HEA Health
- HAL Humanities and Law
- MKD Māori Knowledge and Development
- MIS Mathematical and Information Sciences and Technology
- MED Medicine and Public Health
- PHY Physical Sciences
- SOC Social Sciences and Other Cultural/Social Studies

Notes on Statistics

790 In many cases valid chi square results could not be obtained hence these results comprise percentages which are listed and compared. Percentages have been rounded to 1 decimal place except where cumulative percentages were used when agree/strongly agree and disagree/strongly disagree were combined. The cumulative percentages were then rounded to 1 decimal place.

Response Rate

791 The population for this evaluation was made up of the 168 panellists who participated in the 2003 Quality Evaluation. Of these, 140 responded which equals a response rate of 83.3%.

Notes on Terms used in the Survey

792 When we designed and carried out this survey we used the term 'grade' instead of Quality Category, and 'score' instead of 'rating.' We used the language of the people we spoke to - the people in the scoping interviews, peer review panels and TEOs. This usage did not always align with the *Guidelines*.



Part A: Close-ended (Quantitative) Results of the On-line Survey

Section 1: The Performance-Based Research Fund Model

793 In this section respondents were asked some general questions about the PBRF model. Of particular interest were levels of confidence in the model in principle versus in practice, whether it was a sound funding mechanism and whether they believed PBRF would actually raise the quality of research.

Q1.1 The design of the PBRF quality evaluation framework, in principle, allows for valid assessments of research quality to be made.

794 Three quarters (n=105, 75.0%) of the respondents agreed that the design of the PBRF quality evaluation framework, in principle, allows for valid assessments of research quality to be made and a further 23 (16.4%) strongly agreed. **This resulted in a combined total of those respondents in agreement of 91.4% (n=128).** While 8 respondents (5.7%) were neutral, 4 respondents (2.9%) disagreed.

Table 1: Frequency table for Q1.1 - The design of the PBRF quality evaluation framework, in principle, allows for valid assessments of research quality to be made.

	Frequency	Percent	Cumulative Percent
Strongly agree	23	16.4	16.4
Agree	105	75.0	91.4
Neither	8	5.7	97.1
Disagree	4	2.9	100.0
St.Agree/Agree	128	91.4	
Total	140	100.0	

795 All respondents from the BEC, EDU, ETA and HAL panels either agreed or strongly agreed that the design of the PBRF quality evaluation framework, in principle, allows for valid assessments of research quality to be made This compared with only 66.7% of those on the MKD panel.

796 Valid chi square results could not be obtained but cross-tabulations showed all 100% of chairs agreed or strongly agreed as did 90.6% of non-chairs.

Q1.2 Based on my experience of the implementation of the PBRF quality evaluation framework, in practice, valid assessments of research quality can be made.

797 A little over two thirds of the respondents (n=96, 68.6%) agreed that based on their experience of the implementation of the PBRF quality evaluation framework, in practice, valid assessments of research quality could be made, and a further 28 of those surveyed (20.0%) strongly agreed. **The combined total of those either agreeing or strongly agreeing was 88.6% (n=124).** A combined total of 5.0% (n=7) of respondents either disagreed or strongly disagreed.

Table 2: Frequency table for Q1.2 - Based on my experience of the implementation of the PBRF quality evaluation framework, in practice, valid assessments of research quality can be made.

	Frequency	Percent	Cumulative Percent
Strongly agree	28	20.0	20.0
Agree	96	68.6	88.6
Neither	9	6.4	95.0
Disagree	6	4.3	99.3
St.Agree/Agree	124	88.6	
St.Disagree/Disagree	7	5.0	
Missing	1	0.7	
Total	140	100.0	

798 All of those surveyed from the CPA, EDU, HAL and MED panels agreed or strongly agreed with the statement 'Based on my experience of the implementation of the PBRF quality evaluation framework, in practice, valid assessments of research quality can be made'. Conversely, only 66.7% of those from the ETA panel did so.

799 Cross-tabulations showed that all chairs agreed/strongly agreed compared with 87.5% of non-chairs.

Q1.3 The PBRF model is, in principle, a sound funding mechanism.

800 Compared with questions 1.1 and 1.2, there was less consensus among those surveyed regarding the soundness of the PBRF as a funding mechanism. Less than half of the survey respondents (n=60, 42.9%) agreed with the statement 'The PBRF model is, in principle, a sound funding mechanism' but a further 16.4% (n=23) strongly agreed. In contrast, 12.1% (n=17) disagreed and a further 1.4% (n=2) strongly disagreed. Of note is the quarter (n=37, 26.4%) of respondents who neither agreed nor disagreed.

Table 3: Frequency table for Q1.3 - The PBRF model is, in principle, a sound funding mechanism.

	Frequency	Percent	Cumulative Percent
Strongly agree	23	16.4	16.5
Agree	60	42.9	59.7
Neither	37	26.4	86.3
Disagree	17	12.1	98.6
Strongly Disagree	2	1.4	99.3
St.Agree/Agree	83	59.7	
St.Disagree/Disagree	19	13.5	
Missing	1	0.7	
Total	140	100.0	

801 Higher levels of uncertainty were reflected in the cross-tabulations analysis where the highest percentage of those agreeing that the PBRF model is a sound funding mechanism in principle came from the MED panel (81.8%). Only 33.3% of those from MKD and 36.4% from ETA agreed.

802 Cross-tabulations showed 83.3% of chairs agreed/strongly agreed compared with a much lower figure of only 57.5% for non-chairs.

Q1.4 Assessing the Evidence Portfolios (EPs) of individuals, in principle, will encourage improved research performance.

803 Of the respondents, just over half (51.4%, n=72) agreed with the statement 'Assessing the Evidence Portfolios (EPs) of individuals, in principle, will encourage improved research performance' with a further 22.1% (n=31) strongly agreeing. Approximately three quarters of the respondents (73.6%, n=103) either agreed or strongly agreed. A small combined total of 7.1% (n=10) either disagreed or strongly disagreed and almost one fifth of those surveyed were neutral.



Table 4: Frequency table for Q1.4- Assessing the Evidence Portfolios (EPs) of individuals, in principle, will encourage improved research performance.

	Frequency	Percent	Cumulative Percent
Strongly agree	31	22.1	22.1
Agree	72	51.4	73.6
Neither	27	19.3	92.9
Disagree	8	5.7	98.6
Strongly Disagree	2	1.4	100.0
St.Agree/Agree	103	73.6	
St.Disagree/Disagree	10	7.1	
Total	140	100.0	

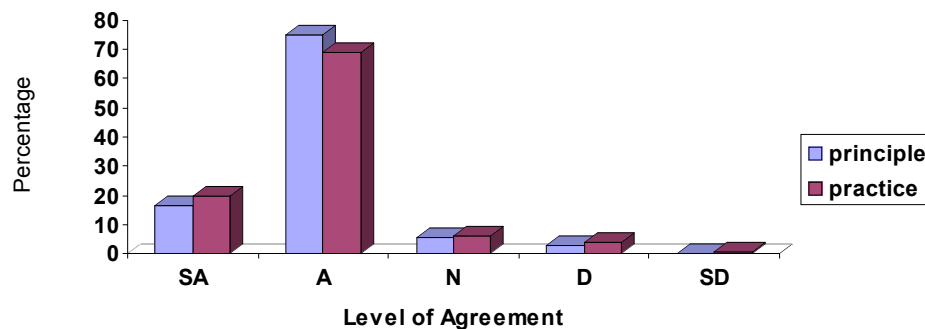
804 The highest percentage of those agreeing that assessing the Evidence Portfolios of individuals, in principle, will encourage improved research performance came from the SOC panel (86.7%). This can be compared with 50.0% of those from the CPA panel.

805 Cross-Cross tabulations showed 83.3% of chairs agreed/strongly agreed compared with 72.7% of non-chairs.

The Performance Based Research Fund Model Section Summary

806 The survey results indicate that most respondents (approximately 90%) either agreed or strongly agreed that the PBRF quality evaluation framework allows for valid assessments of research quality to be made both in principle and in practice. Moreover, the correlation analysis suggests a fairly strong positive relationship between these two variables (Q1.1 and 1.2) where those who agreed with the former also agreed with the latter.

Figure 1: Comparison of agreement between Q1.1 (Principle) and Q1.2 (Practice).





- 807 All respondents from the BEC, EDU, ETA and HAL panels agreed that the PBRF worked in principle compared with two thirds of those from the MKD panel. A similar, but not identical, result was obtained regarding the PBRF in practice where all those from EDU, HAL, CPA and MED panels agreed. Only two thirds of those from the BEC panel agreed.
- 808 **Lower levels of agreement** were found regarding the **soundness of the PBRF model as a funding mechanism**. A higher level of uncertainty is evident in the quarter of respondents who neither agreed nor disagreed with the statement in Q 1.3, however, it must be pointed out that the majority of the remainder (59.7%) either agreed or strongly agreed that the PBRF was sound in principle. The highest percentage of those agreeing that the PBRF model is a sound funding mechanism in principle came from the MED and the lowest from MKD. Cross-tabulations showed **83.3% of chairs** agreed/strongly agreed compared with only **57.5% of non-chairs**.
- 809 Almost **three quarters of those surveyed agreed to some extent that assessing the Evidence Portfolios of individuals would, in principle, encourage improved research performance** and most of the remainder neither agreed nor disagreed. The highest percentage of those in agreement came from the SOC panel and the lowest from CPA.

Section 2: Indicators of Research Quality and the Quality Categories

810 This section explored the respondents' views on a range of issues regarding the indicators of research quality (NROs, PE, CRE) and the Quality Categories (A, B, C and R).

Q2.1 Nominated Research Outputs (NROs) are valid indicators of research quality.

811 Of the respondents, 60.0% (n=84) agreed with the statement 'Nominated Research Outputs (NROs) are valid indicators of research quality' and a further 32.1% (n=45) strongly agreed making a comparatively high total of **92.8% of respondents in agreement.**

Table 5: Frequency table for Q2.1 - Nominated Research Outputs (NROs) are valid indicators of research quality.

	Frequency	Percent	Cumulative Percent
Strongly agree	45	32.1	32.4
Agree	84	60.0	92.8
Neither	6	4.3	97.1
Disagree	4	2.9	99.3
St.Agree/Agree	129	92.8	
Missing	1	0.7	
Total	140	100.0	

812 All members of the BS, HAL, HEA and MIS panels agreed that Nominated Research Outputs are valid indicators of research quality. This compared with the lower percentage of 81.8% from the PHY panel.

813 Cross-tabulations showed 100% of chairs agreed/strongly agreed compared with 92.2% of non-chairs.

Q2.2 Peer Esteem (PE) is a valid indicator of research quality.

814 Almost two thirds (63.6%, n=89) of respondents agreed that Peer Esteem is a valid indicator of research quality and a further 15.7% (n=22) strongly agreed. This resulted in a combined total of just over three quarters (79.3%). The number of respondents who were either neutral (11.4%, n=16) or disagreed (9.3%, n=13) were roughly similar.



Table 6: Frequency table for Q2.2 - Peer Esteem (PE) is a valid indicator of research quality.

	Frequency	Percent	Cumulative Percent
Strongly agree	22	15.7	15.7
Agree	89	63.6	79.3
Neither	16	11.4	90.7
Disagree	13	9.3	100.0
St.Agree/Agree	111	79.3	
Total	140	100.0	

815 All members of the MIS and MKD panels agreed that Peer Esteem is a valid indicator of research quality, compared with only 53.8% of those from BEC.

816 Cross-tabulations showed 91.7% of chairs agreed/strongly agreed compared with 78.1% of non-chairs.

Q2.3 Contribution to the Research Environment (CRE) is a valid indicator of research quality.

817 **Slightly less than three quarters** of the respondents **either agreed** (57.9%, n=81) **or strongly agreed** (13.6%, n=19) that 'Contribution to the Research Environment (CRE) is a valid indicator of research quality'. Slightly greater numbers were neutral (17.9%, n=25) than disagreed (10.7%, n=15).

Table 7: Frequency table for Q2.3 - Contribution to the Research Environment (CRE) is a valid indicator of research quality.

	Frequency	Percent	Cumulative Percent
Strongly agree	19	13.6	13.6
Agree	81	57.9	71.4
Neither	25	17.9	89.3
Disagree	15	10.7	100.0
St.Agree/Agree	100	71.4	
Total	140	100.0	

818 All members of the EDU panel agreed with the statement 'Contribution to the Research Environment (CRE) is a valid indicator of research quality' compared with 50.0% of those from CPA.



819 Cross-tabulations showed 83.3% of chairs agreed/strongly agreed compared with 70.3% of non-chairs.

Q2.4 The combination of NROs, PE and CRE is a valid indicator of research quality.

820 Slightly over half of the respondents (54.3%, n=76) agreed with the statement that 'The combination of NROs, PE and CRE is a valid indicator of research quality' with an additional 25.0% (n=35) strongly agreeing. This resulted in a combined total of 81.0% (n=111) indicating some level of agreement. A combined total of 15.0% (n=11) were either neutral or did not answer the question and 5.7% (n=8) of those surveyed disagreed.

Table 8: Frequency table for Q2.4 - The combination of NROs, PE and CRE is a valid indicator of research quality.

	Frequency	Percent	Cumulative Percent
Strongly agree	35	25.0	25.0
Agree	76	54.3	81.0
Neither	18	12.9	94.2
Disagree	8	5.7	97.9
St.Agree/Agree	111	81.0	
Missing	3	2.1	
Total	140	100.0	

821 High numbers of members from most panels agreed that the combination of NROs, PE and CRE is a valid indicator of research quality, compared with 60.0% of those on the CPA panel.

822 Cross-tabulations showed 72.7% of chairs agreed/strongly agreed which was less than the 81.7% of non-chairs who did so.

Q2.5 The weighting between NROs, PE and CRE (70:15:15) is about right.

823 Just over half of the respondents (52.1%, n=73) agreed with the statement 'The weighting between NROs, PE and CRE (70:15:15) is about right' and a further 10.7% (n=15) strongly agreed. This resulted in a **combined total of approximately two thirds indicating agreement**. Approximately equal numbers were either neutral (17.1%, n=24) or disagreed (17.9%, n=25). Only two respondents (1.4%) strongly disagreed.



Table 9: Frequency table for Q2.5 - The weighting between NROs, PE and CRE (70:15:15) is about right.

	Frequency	Percent	Cumulative Percent
Strongly agree	15	10.7	10.8
Agree	73	52.1	63.3
Neither	24	17.1	80.6
Disagree	25	17.9	98.6
Strongly Disagree	2	1.4	99.3
St.Agree/Agree	88	63.3	
St.Disagree/Disagree	27	19.3	
Missing	1	0.7	
Total	140	100.0	

824 The panel with the highest percentage of members agreeing with the statement 'The weighting between NROs, PE and CRE (70:15:15) is about right' was EDU at 87.5%. This compared with only 50.0% of those on the CPA panel.

825 Cross-tabulations showed 66.7% of chairs agreed/strongly agreed compared with an almost equal percentage (63.0%) of non-chairs.

Q2.6 It is appropriate to translate scores into grades.

826 Less than half of those surveyed (45.0%, n=63) **agreed** that it is appropriate to translate scores into grades, however, an additional 12.9% (n=18) **strongly agreed** making a **combined total of only 58.3%** (n=81). Approximately one fifth (21.4%, n=30) were neutral and a combined total of 20.0% (n=28) either disagreed (17.1%, n=24) or strongly disagreed (2.9%, n=4).



Table 10: Frequency table for Q 2.6 - It is appropriate to translate scores into grades.

	Frequency	Percent	Cumulative Percent
Strongly agree	18	12.9	12.9
Agree	63	45.0	58.3
Neither	30	21.4	79.9
Disagree	24	17.1	97.1
Strongly Disagree	4	2.9	99.3
St.Agree/Agree	81	58.3	
St.Disagree/Disagree	28	20.0	
Missing	1	0.7	
Total	140	100.0	

827 All members of the CPA panel agreed that it is appropriate to translate scores into grades, however, only 33.0% of those from the ETA panel did so.

828 Cross-tabulations showed **83.3% of chairs** agreed/strongly agreed compared with **only 55.9% of non-chairs**.

Q2.7 Using the letter grades (A B C R) resulted in an adequate differentiation of the range of research quality in the EPs submitted.

829 **Almost equal numbers** of the respondents **agreed** (35.0%, n=49) **and disagreed** (33.6%, n=47) with the statement 'Using the letter grades (A B C R) resulted in an adequate differentiation of the range of research quality in the EPs submitted'. Only 5.7% (n=8) strongly agreed compared with 15.7% (n=22) who strongly disagreed.

Table 11: Frequency table for Q 2.7 - Using the letter grades (A B C R) resulted in an adequate differentiation of the range of research quality in the EPs submitted.

	Frequency	Percent	Cumulative Percent
Strongly agree	8	5.7	5.7
Agree	49	35.0	40.7
Neither	14	10.0	50.7
Disagree	47	33.6	84.3
Strongly Disagree	22	15.7	100.0
St.Agree/Agree	57	40.7	
St.Disagree/Disagree	69	49.3	
Total	140	100.0	

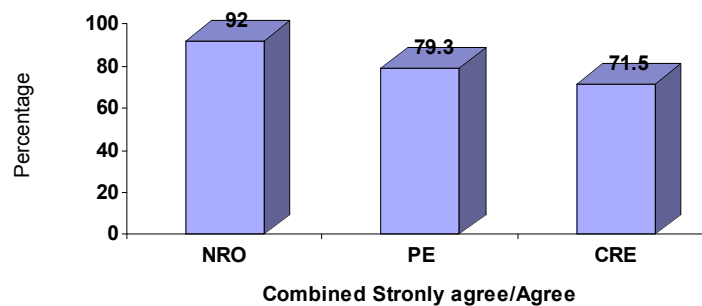
830 At 69.2%, HEA was the panel with the highest percentage of members agreeing with the statement 'Using the letter grades (A B C R) resulted in an adequate differentiation of the range of research quality in the EPs submitted'. The lowest percentage (16.7%) came from CPA.

831 Cross-tabulations showed only half of the chairs agreed/strongly agreed but this was still higher than the 39.8% of non-chairs who agreed.

Indicators of Research Quality and Quality Categories Section Summary

832 Although there were variations in the ratio of strongly agree and agree responses to each of the statements regarding the three quality indicators (NROs, PE and CRE), the figure below shows that, overall, **a higher percentage of respondents agreed (combined – 92.8%) that Nominated Research Outputs are valid indicators of research quality.**

Figure 2: Agreement over the Validity of the Three Quality Indicators.



833 A combined total of 81.0% indicated agreement with the statement 'The combination of NROs, PE and CRE is a valid indicator of research



quality'. A combined total of 15.0% (n=11) were either neutral or did not answer the question. A slightly lower combined agreement figure of about two thirds (63.3%) was found for the statement 'The weighting between NROs, PE and CRE (70:15:15) is about right'. Approximately equal numbers were either neutral (17.1%, n=24) or disagreed (17.9%, n=25). The statement 'It is appropriate to translate scores into grades' received even lower levels of combined agreement (58.3%, n=81). Almost even numbers were either neutral or disagreed.

- 834 The statement 'Using the letter grades (A B C R) resulted in an adequate differentiation of the range of research quality in the EPs submitted' is notable in that **almost equal numbers** of the respondents **agreed** (35.0%, n=49) **and disagreed** (33.6%, n=47). Furthermore only 5.7% (n=8) strongly agreed compared with 15.7% (n=22) who strongly disagreed. This points to an area where further investigations are required⁷.
- 835 The **CPA panel** had **the lowest percentage** of members **agreeing to four of the seven statements** in this section (2.3 - CRE is a valid indicator; 2.4 -The combination of NROs, PE and CRE is a valid indicator; 2.5- The weighting between NRO, PE and CRE is about right and; 2.7- The use of letter grades resulted in adequate differentiation).
- 836 The statement in question 2.4 'The combination of NROs, PE and CRE is a valid indicator of research quality' was the only one where a higher percentage of non-chairs (81.7%) agreed/strongly agreed than did chairs (72.7%). Also of interest were the results of the cross-tabulations of Q 2.6 (regarding the translation of scores into grades) which showed **83.3% of chairs** agreed/strongly agreed compared with **only 55.9% of non-chairs**.

⁷ See the Qualitative Results in Part 2.

Section 3: The Impact of the Quality Evaluation of Individual's Evidence Portfolios

837 In this section, respondents were asked their views on a range of issues concerning the evidence portfolios, such as their validity as indicators of research quality both within and between disciplines, whether they provided a fair measure of research quality, and whether the PBRF is likely to encourage more individual or collaborative research efforts.

Q3.1 Evidence Portfolios (EPs) provide a valid measure of individual research performance.

838 Two thirds (67.9%, n=95) of those surveyed agreed that Evidence Portfolios (EPs) provide a valid measure of individual research performance and an additional 22.1% (n=31) strongly agreed. This resulted in a **combined total of 90.0%** (n=126) indicating some level of **agreement**.

Table 12: Frequency table for Q 3.1 - Evidence Portfolios (EPs) provide a valid measure of individual research performance.

	Frequency	Percent	Cumulative Percent
Strongly agree	31	22.1	22.1
Agree	95	67.9	90.0
Neither	11	7.9	97.9
Disagree	3	2.1	100.0
St.Agree/Agree	126	90.0	
Total	140	100.0	

839 All members of the BEC, EDU and MIS panels agreed that Evidence Portfolios provide a valid measure of individual research performance, compared with 66.7% of those from the MKD panel.

840 Cross-tabulations showed **all 100% of chairs** agreed/strongly agreed compared with **89.1%** of non-chairs.

Q3.2 EPs allowed valid comparisons of individual research performance to be made within the same discipline.

841 Two thirds (66.4%, n=93) of the respondents agreed with the statement 'EPs allow valid comparisons of individual research performance to be made within the same discipline' and another 26.4% (n=37) strongly agreed. This resulted in a combined total of 92.9% with the remainder either neutral (2.9%, n=4) or disagreeing (4.3%, n=6).

Table 13: Frequency table for Q 3.2 - EPs allow valid comparisons of individual research performance to be made within the same discipline.

	Frequency	Percent	Cumulative Percent
Strongly agree	37	26.4	26.4
Agree	93	66.4	92.9
Neither	4	2.9	95.7
Disagree	6	4.3	100.0
St.Agree/Agree	130	92.9	
Total	140	100.0	

842 All members of the BS, ETA, HAL, MED and MIS panels agreed that EPs allow valid comparisons of individual research performance to be made within the same discipline, however, only 66.7% of those from MKD agreed.

843 Cross-tabulations showed **all 100% of chairs** agreed/strongly agreed compared with **only 42.2%** of non-chairs.

Q3.3 EPs allow valid comparisons of individual research performance to be made across different disciplines.

844 Only 40.7% (n=57) of the respondents agreed that EPs allow valid comparisons of individual research performance to be made across different disciplines, with another 3.6% (n=5) strongly agreeing. This resulted in a low combined agreement of 44.3% (n=62). Similar numbers were either neutral (27.1%, n=38) or disagreed (25.7%, n=36), and 2.9% (n=4) strongly disagreed.

Table 14: Frequency table for Q 3.3 - EPs allow valid comparisons of individual research performance to be made across different disciplines.

	Frequency	Percent	Cumulative Percent
Strongly agree	5	3.6	3.6
Agree	57	40.7	44.3
Neither	38	27.1	71.4
Disagree	36	25.7	97.1
Strongly Disagree	4	2.9	100.0
St.Agree/Agree	62	44.3	
St.Disagree/Disagree	40	28.6	
Total	140	100.0	

- 845 Higher numbers from the SOC panel agreed that EPs allow valid comparisons of individual research performance to be made across different disciplines, compared with only 16.7% of those from MKD.
- 846 Cross-tabulations showed **three quarters (75.0%) of chairs** agreed/strongly agreed compared with **only 41.4%** of non-chairs.

Q3.4 Processes associated with the quality evaluation of EPs were consistent within the same discipline.

- 847 Although only 55.0% (n=77) of the respondents **agreed** with the statement 'Processes associated with the quality evaluations of EPs were consistent within the same discipline', another 30.7% (n=43) **strongly agreed** resulting in a **combined total of 85.7%** (n=120). Equal numbers (7.1%, n=10) were either neutral or disagreed.

Table 15: Frequency table for Q 3.4 - Processes associated with the quality evaluations of EPs were consistent within the same discipline.

	Frequency	Percent	Cumulative Percent
Strongly agree	43	30.7	30.7
Agree	77	55.0	85.7
Neither	10	7.1	92.9
Disagree	10	7.1	100.0
St.Agree/Agree	120	85.7	
Total	140	100.0	

- 848 All those from HEA and MIS agreed with the statement 'Processes associated with the quality evaluations of EPs were consistent within the same discipline'. However, only 61.5% of members from BEC did so.
- 849 Cross-tabulations showed **all 100% of chairs** agreed/strongly agreed compared with **84.4%** of non-chairs.

Q3.5 Processes associated with the quality evaluation of EPs were consistent across different disciplines.

- 850 Almost half of those surveyed (46.4%, n=65) neither agreed nor disagreed with the statement 'Processes associated with the quality evaluations of EPs were consistent across different disciplines'. Approximately one third (32.8%, n=46) either agreed (27.1%, n=38) or strongly agreed (5.7%, n=8), but a combined total of 20.0% (n=28) either disagreed (17.9%, n=25) or strongly disagreed (2.1%, n=3). In summary, one third (33.1%, n=46) indicated some level of agreement.

Table 16: Frequency table for Q 3.5 - Processes associated with the quality evaluations of EPs were consistent across different disciplines.

	Frequency	Percent	Cumulative Percent
Strongly agree	8	5.7	5.8
Agree	38	27.1	33.1
Neither	65	46.4	79.9
Disagree	25	17.9	97.8
Strongly Disagree	3	2.1	99.3
St.Agree/Agree	46	33.1	
St.Disagree/Disagree	28	20.0	
Missing	1	0.7	
Total	140	100.0	

851 The highest percentage (58.3%) of those who agreed that processes associated with the quality evaluations of EPs were consistent across different disciplines came from the HEA panel. The lowest percentage (15.4%) of those in agreement with this statement came from the BEC panel.

852 Cross-tabulations showed **two thirds (66.7%) of chairs** agreed/strongly agreed compared with **only 29.9%** of non-chairs.

Q3.6 The individual results of the quality evaluation are unlikely to be used in making decisions about the terms or conditions of employment of researchers.

853 Almost **one third** of those surveyed (62.8%, n=88) either **disagreed** (42.1%, n=59) or **strongly disagreed** (20.7%, n=29) with the statement 'The individual results of the quality evaluation are unlikely to be used in making decisions about the terms or conditions of employment of researchers'. While approximately one quarter of those surveyed (26.4%, n=37) were neutral, only 7.1% (n=10) agreed or strongly agreed (2.1%, n=3).



Table 17: Frequency table for Q 3.6 - The individual results of the quality evaluation are unlikely to be used in making decisions about the terms or conditions of employment of researchers.

	Frequency	Percent	Cumulative Percent
Strongly agree	3	2.1	2.2
Agree	10	7.1	9.4
Neither	37	26.4	36.2
Disagree	59	42.1	79.0
Strongly Disagree	29	20.7	98.6
St.Agree/Agree	13	9.4	
St.Disagree/Disagree	88	62.8	
Missing	2	1.4	
Total	140	100.0	

854 Comparatively low percentage figures were found for this variable. Only one third of members from the HEA panel agreed that the individual results of the quality evaluation are unlikely to be used in making decisions about the terms or conditions of employment of researchers. No one from CPA, MIS or SOC agreed.

855 Cross-tabulations showed only 9.1% of chairs agreed/strongly agreed and as did only 9.4% of non-chairs. A higher percentage of **chairs (63.5%** compared with 23.6% of non-chairs) were **neutral** and over **two third of non-chairs disagreed/strongly disagreed.**

Q3.7 The PBRF is likely to promote more individual research efforts than prior to the PBRF.

856 Over half of the respondents (54.3%, n=76) agreed that the PBRF is likely to promote more individual research efforts than prior to the PBRF, and an additional 21.4% (n=30) strongly agreed. Only 7.1% (n=10) disagreed with this statement.

Table 18: Frequency table for Q 3.7 - The PBRF is likely to promote more individual research efforts than prior to the PBRF.

	Frequency	Percent	Cumulative Percent
Strongly agree	30	21.4	21.6
Agree	76	54.3	76.3
Neither	22	15.7	92.1
Disagree	10	7.1	99.3
Strongly Disagree	1	0.7	
St.Agree/Agree	106	76.3	
St.Disagree/Disagree	11	7.8	
Missing	1	0.7	
Total	140	100.0	

857 All members of EDU and MIS agreed with the statement 'The PBRF is likely to promote more individual research efforts than prior to the PBRF', however, only half those from BEC agreed.

858 Cross-tabulations showed almost equal numbers of chairs (75.0%) and non-chairs (76.4%) agreed/strongly agreed with this statement.

3.8 The PBRF is likely to promote more collaborative research efforts than prior to the PBRF.

859 Over one third of the respondents (39.3%, n=55) neither agreed nor disagreed that the PBRF is likely to promote more collaborative research efforts than prior to the PBRF. Another third (34.3%, n=48) either agreed (28.6%, n=40) or strongly agreed (5.7%, n=8), and 19.3% (n=27) disagreed or strongly disagreed (7.1%, n=10).

Table 19: Frequency table for Q 3.8 - The PBRF is likely to promote more collaborative research efforts than prior to the PBRF.

	Frequency	Percent	Cumulative Percent
Strongly agree	8	5.7	5.7
Agree	40	28.6	34.3
Neither	55	39.3	73.6
Disagree	27	19.3	92.9
Strongly Disagree	10	7.1	100.0
St.Agree/Agree	48	34.3	
St.Disagree/Disagree	37	26.4	
Total	140	100.0	



860 The highest percentage (53.8%) of those agreeing with the statement 'The PBRF is likely to promote more collaborative research efforts than prior to the PBRF' were found for the HEA panel and the lowest percentage (8.3%) for ETA.

861 Cross-tabulations showed only 25.0 of chairs agreed/strongly agreed compared with a slightly higher percentage (35.2%) of non-chairs.

Q3.9 The PBRF is likely to ensure that research funding is concentrated where research is of the highest quality.

862 Half of those surveyed (50.7%, n=71) **agreed** with the statement that 'The PBRF is likely to ensure that research funding is concentrated where research is of the highest quality' and a further 14.3% (n=20) **strongly agreed**. This resulted in a **combined total of 65.5%** (n=91). A quarter of the respondents neither agreed nor disagreed (25.0%, n=35), and 6.4% (n=9) disagreed or strongly disagreed (2.9%, n=4).

Table 20: Frequency table for Q 3.9 - The PBRF is likely to ensure that research funding is concentrated where research is of the highest quality.

	Frequency	Percent	Cumulative Percent
Strongly agree	20	14.3	14.4
Agree	71	50.7	65.5
Neither	35	25.0	90.6
Disagree	9	6.4	97.1
Strongly Disagree	4	2.9	99.3
St.Agree/Agree	83	65.5	
St.Disagree/Disagree	19	9.3	
Missing	1	0.7	
Total	140	100.0	

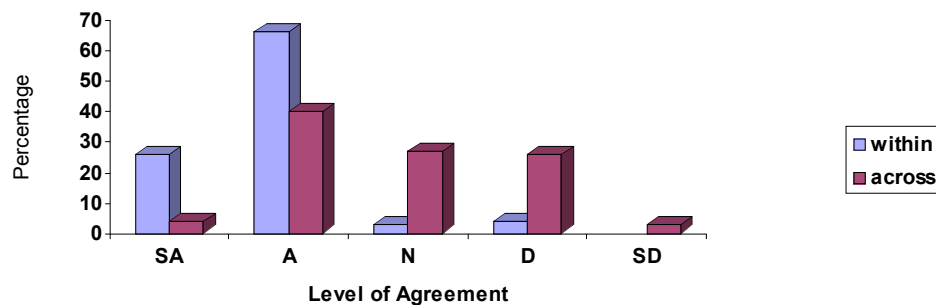
863 The highest percentages of those agreeing that the PBRF is likely to ensure that research funding is concentrated where research is of the highest quality were found for the ETA and MED panels. This compared with only 30.8% of those from BEC.

864 Cross-tabulations showed 83.3% of chairs agreed/strongly agreed compared with only 63.8% of non-chairs.

The Impact of the Quality Evaluation of Individual's Evidence Portfolios Section Summary

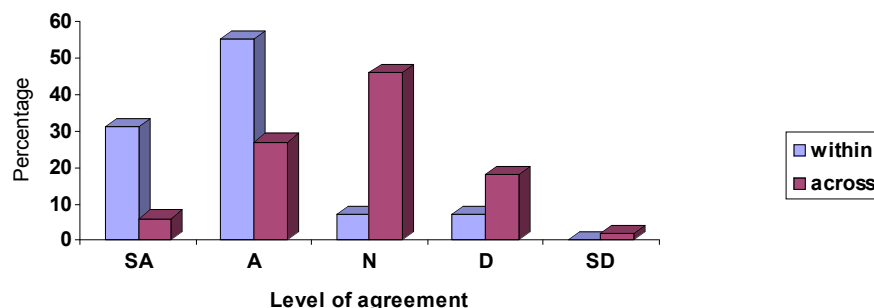
865 While there was a **very high level of agreement** that Evidence Portfolios provide a **valid measure of individual research performance** (90.0%) and that Evidence Portfolios allow valid comparisons of individual research performance to be made **within the same discipline** (92.9%), there was **far less agreement** that Evidence Portfolios allow comparisons of individual research performance to be made **across** different disciplines (44.3%) as illustrated in the figure below.

Figure 3: Agreement over Validity of Comparisons of Performance Within and Across Disciplines.



866 A similar pattern emerged during analysis of levels of agreement with the statements 'Processes associated with the quality evaluations of EPs were consistent **within the same discipline**' (**85.7%**) and 'Processes associated with the quality evaluations of EPs were consistent **across different disciplines**' (**33.1%**) as illustrated in the figure below. This points to an area where further research should be directed⁸.

Figure 4: Agreement over Consistency of Processes Within and Across Disciplines.

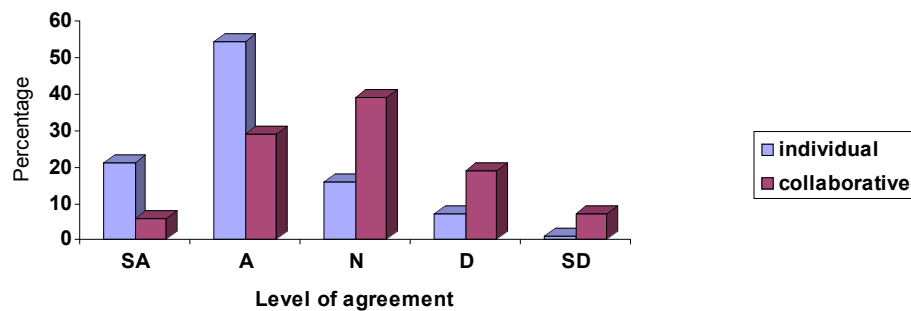


⁸ See the Qualitative Results, Part 2.

867 Responses to the statement 'The individual results of the quality evaluation are unlikely to be used in making decisions about the terms or conditions of employment of researchers' are notable because **the number of those disagreeing (62.8%, n=88) outweighed those agreeing (9.4%, n=13)**. While this particular question does not establish whether the likely impact is positive or negative, the qualitative results certainly suggest various expectations of generally adverse effects⁹.

868 While **three quarters (75.7%)** of those surveyed **indicated agreement** with the statement 'The PBRF is likely to promote more **individual research** efforts than prior to the PBRF' there was greater ambivalence in responses to the statement 'The PBRF is likely to promote more **collaborative research** efforts than prior to the PBRF'. Only one third (34.3%) indicated agreement and **over one third (39.3%) were unsure** as illustrated in the figure below.

Figure 5: Agreement that the PBRF will Promote Individual and Collaborative Research Efforts.



869 Although over two thirds (65.5%) indicated agreement with the statement 'The PBRF is likely to ensure that research funding is concentrated where research is of the highest quality', a degree of uncertainty is evident in the one quarter of respondents who neither agreed nor disagreed.

870 The **HEA panel had the highest percentage of members agreeing** with four of the nine statements in this section (3.4 – processes were consistent within the same discipline; 3.5- processes were consistent across disciplines; 3.6 – individual results are unlikely to be used and; 3.8 PBRF will promote collaborative efforts). The **BEC panel had the lowest percentage of members agreeing** with three of the nine statements in this section (3.4 –processes were consistent within disciplines; 3.7 PBRF will promote individual research efforts and; 3.9 PBRF will ensure funding is concentrated where research is of highest quality).

⁹ See the Qualitative Results, Part 2.



871 Cross-tabulations showed that, in general, a higher percentage of chairs agreed with statements in this section than did non-chairs (with the exceptions of 3.6, 3.7 and 3.8 which were fairly close). Of note were the responses to Q3.2 (valid comparisons can be made within the same discipline) and Q3.5 (processes were consistent within disciplines) where the percentage of chairs agreeing was substantially more than (a difference of more than 35%) that of non-chairs. While similarly low percentages agreed (approximately 9%) that the individual results were unlikely to be used in decisions about the terms or conditions of employment, chairs were more likely to be neutral (63.6% compared with 23.6% of non-chairs) and non-chairs were more likely to disagree (66.9% compared with 27.3% of chairs).

Section 4: The Operation of the Peer Review Panel

872 This section investigated the respondents' views on the operation of the peer review panels, their processes and composition, and the forms of support panellists received. This addressed questions about whether the respondents thought the final Quality Categories awarded were fair and consistent.

Q4.1 Overall, the peer review panel process was sound.

873 A little over half of those surveyed (50.7%, n=71) agreed that overall, the peer review panel process was sound and an additional 42.1% (n=59) strongly agreed. This resulted in a combined total of 92.9%.

Table 21: Frequency table for Q 4.1 - Overall, the peer review panel process was sound.

	Frequency	Percent	Cumulative Percent
Strongly agree	59	42.1	42.1
Agree	71	50.7	92.9
Neither	6	4.3	97.1
Disagree	4	2.9	100.0
St.Agree/Agree	130	92.9	
Total	140	100.0	

874 All members of the BS, CPA, EDU, ETA, HEA, and MIS panels agreed with the statement 'Overall, the peer review panel process was sound'. Only 76.9% of those from the BEC panel agreed. The mean agreement for this variable was 93.6%.

875 Cross-tabulations showed all 100% of chairs agreed/strongly agreed compared with 92.2% of non-chairs.

Q4.2 The composition of my peer review panel ensured that we made balanced judgements.

876 Although only a little over half of those surveyed agreed with the statement 'The composition of my peer review panel ensured that we made balanced judgements', a further 37.1% (n=52) strongly agreed bringing the **combined total to 88.6%** (n=124). Most of the remaining respondents (8.6%, n=8) neither agreed nor disagreed.

Table 22: Frequency table for Q 4.2 - The composition of my peer review panel ensured that we made balanced judgements.

	Frequency	Percent	Cumulative Percent
Strongly agree	52	37.1	37.1
Agree	72	51.4	88.6
Neither	12	8.6	97.1
Disagree	4	2.9	100.0
St.Agree/Agree	124	88.6	
Total	140	100.0	

877

878 All members of the HEA, MIS and MKD panels agreed that the composition of my peer review panel ensured that we made balanced judgements but only 61.5% of those from BEC agreed.

879 Cross-tabulations showed all 100% of chairs agreed/strongly agreed compared with 87.5% of non-chairs.

Q4.3 The Panel Member's Deskfile was a useful tool in preparing me for my role.

880 Half those surveyed (50.6%, n=71) agreed that The Panel Member's Deskfile was a useful tool in preparing them for their role and another 21.4% (n=30) strongly agreed. In contrast, 10.0% (n=14) of the respondents disagreed and 2.1% (n=3) strongly disagreed.

Table 23: Frequency table for Q 4.3 - The Panel Member's Deskfile was a useful tool in preparing me for my role.

	Frequency	Percent	Cumulative Percent
Strongly agree	30	21.4	21.4
Agree	71	50.7	72.1
Neither	22	15.7	87.9
Disagree	14	10.0	97.9
Strongly Disagree	3	2.1	100.0
St.Agree/Agree	101	72.1	
St.Disagree/Disagree	17	12.1	
Total	140	100.0	

881 All those from the EDU panel agreed with the statement that 'The Panel Member's Deskfile was a useful tool in preparing me for my role', however, only 33.3% of those from the CPA panel did so.

882 **Chairs appeared less impressed with the Deskfile than non-chairs.** Cross-tabulations showed only 66.7% of chairs agreed/strongly agreed compared with nearly all (97.7%) of non-chairs.

Q4.4 The TEC secretariat supported me and my panel effectively.

883 The **clear majority** (62.9% n =88) of the respondents **strongly agreed** with the statement 'The TEC secretariat supported me and my panel effectively' and a further 35.0% (n=49) agreed. This resulted in a combined total of 97.9% (n=137).

Table 24: Frequency table for Q 4.4 - The TEC secretariat supported me and my panel effectively.

	Frequency	Percent	Cumulative Percent
Strongly agree	88	62.9	62.9
Agree	49	35.0	97.9
Neither	2	1.4	99.3
Disagree	1	0.7	100.0
St.Agree/Agree	137	97.9	
Total	140	100.0	

884 All members from most of the panels agreed that the TEC secretariat supported them and their panel effectively. The exceptions were ETA (91.7%), MED (90.9%) and SOC (93.3%).

885 Cross-tabulations showed all 100% of chairs agreed/strongly agreed and this was close to the similar figure of 97.7% of non-chairs.

Q4.5 The process by which my peer review panel allocated numerical scores was fair.

886 Relatively high numbers of respondents (combined 90.7%, n=127) either agreed (57.1%, n=80) or strongly agreed (33.6%, n=47) that the process by which their peer review panel allocated numerical scores was fair. Of the remainder, 5.7% (n=8) respondents were neutral and 2.9% (n=4) disagreed or strongly disagreed (0.7%, n=1).



Table 25: Frequency table for Q 4.5 - The process by which my peer review panel allocated numerical scores was fair.

	Frequency	Percent	Cumulative Percent
Strongly agree	47	33.6	33.6
Agree	80	57.1	90.7
Neither	8	5.7	96.4
Disagree	4	2.9	99.3
Strongly Disagree	1	0.7	100.0
St.Agree/Agree	127	90.7	
St.Disagree/Disagree	5	3.6	
Total	140	100.0	

887 All members of the CPA, ETA, HEA, and MIS panels agreed with the statement 'The process by which my peer review panel allocated numerical scores was fair'. This compared with only 76.9% of those from the BS panel.

888 Cross-tabulations showed similar proportions of chairs (91.7%) and non-chairs (90.6%) agreed/strongly agreed.

Q4.6 The process by which my peer review panel allocated interim grades was fair.

889 Almost two thirds of the respondents (62.1%, n=87) agreed with the statement 'The process by which my peer review panel allocated interim grades was fair' and a further 27.9% (n=39) strongly agreed. This results in a combined total of 90.6%. A combined total of 6.4% (n=9) were either neutral or did not answer this question. A small number either disagreed (2.1%, n=3) or strongly disagreed (1.4%, n=2).

Table 26: Frequency table for Q 4.6 - The process by which my peer review panel allocated interim grades was fair.

	Frequency	Percent	Cumulative Percent
Strongly agree	39	27.9	28.1
Agree	87	62.1	90.6
Neither	8	5.7	96.4
Disagree	3	2.1	98.6
Strongly Disagree	2	1.4	99.3
St.Agree/Agree	126	90.6	
St.Disagree/Disagree	5	3.5	
Missing	1	0.7	
Total	140	100.0	

890 All members of the HEA and MIS panels agreed that the process by which my peer review panel allocated interim grades was fair. This compared with the 76.9% of those from the BEC panel who agreed.

891 Cross-tabulations showed all 100% of chairs agreed/strongly agreed compared with 89.8% of non-chairs.

Q4.7 The moderation process ensured that the allocation of final grades was consistent across disciplines.

892 Although more respondents (42.1%, n=59) agreed that 'The moderation process ensured that the allocation of final grades was consistent across disciplines' than disagreed (5.0%, n=7), one third of those surveyed (33.6%, n=47) neither agreed nor disagreed.

Table 27: Frequency table for Q 4.7 - The moderation process ensured that the allocation of final grades was consistent across disciplines.

	Frequency	Percent	Cumulative Percent
Strongly agree	23	16.4	16.8
Agree	59	42.1	58.5
Neither	47	33.6	92.1
Disagree	7	5.0	97.1
Strongly Disagree	1	0.7	97.9
St.Agree/Agree	82	59.9	
St.Disagree/Disagree	8	5.7	
Missing	3	2.1	
Total	140	100.0	

- 893 The panel with the highest percentage of those agreeing that the moderation process ensured that the allocation of final grades was consistent across disciplines was MED at 90.9%. The panel with the lowest percentage was MKD at 16.7%.
- 894 Cross-tabulations showed 83.3% of chairs agreed/strongly agreed compared with only 57.6% of non-chairs.

Q4.8 Participating TEOs will accept that the final grades were allocated fairly.

895 A **relatively high level of uncertainty** was evident in responses to the statement 'participating TEOs will accept that the final grades were allocated fairly' with nearly half those surveyed (48.6%, n=68) neither agreeing nor disagreeing. Overall, more respondents agreed (35.0%, n=49) or strongly agreed (3.6%, n=5) than disagreed (9.3%. n=13) or strongly disagreed (2.1%, n=3).

Table 28: Frequency table for Q 4.8 - Participating TEOs will accept that the final grades were allocated fairly.

	Frequency	Percent	Cumulative Percent
Strongly agree	5	3.6	3.6
Agree	49	35.0	39.1
Neither	68	48.6	88.4
Disagree	13	9.3	97.8
Strongly Disagree	3	2.1	98.6
St.Agree/Agree	54	39.1	
St.Disagree/Disagree	16	11.4	
Missing	2	1.4	
Total	140	100.0	

- 896 The panel with the highest percentage of those agreeing with the statement 'Participating TEOs will accept that the final grades were allocated fairly' was PHY at 72.2%. Conversely, the lowest percentages were found among those from MED at 9.1%.
- 897 Low levels of agreement were found for both chairs and non-chairs. Cross-tabulations showed only 45.5% of chairs agreed/strongly agreed and only 38.6% of non-chairs indicated agreement.



Q4.9 Participating researchers will accept that the final grades were allocated fairly.

898 The majority of the respondents (52.9%, n=74) neither agreed nor disagreed with the statement 'Participating researchers will accept that the final grades were allocated fairly'. The remaining answers were fairly evenly split between those who agreed (23.6%, n=33) or strongly agreed (2.1, n=3) and those who either disagreed (15.7%, n=22) or strongly disagreed (4.3, n=6).

Table 29: Frequency table for Q 4.9 - Participating researchers will accept that the final grades were allocated fairly.

	Frequency	Percent	Cumulative Percent
Strongly agree	3	2.1	2.2
Agree	33	23.6	26.1
Neither	74	52.9	79.7
Disagree	22	15.7	95.7
Strongly Disagree	6	4.3	98.6
St.Agree/Agree	36	26.1	
St.Disagree/Disagree	28	20.0	
Missing	2	1.4	
Total	140	100.0	

899 The panel with the highest percentage of members agreeing with the statement 'Participating researchers will accept that the final grades were allocated fairly' was ETA at 41.7% and the lowest was MED at 9.1%.

900 **Low levels of agreement** were found for this statement. Cross-tabulations showed only 36.4% of chairs agreed/strongly agreed and only one quarter (25.2%) of non-chairs did so.

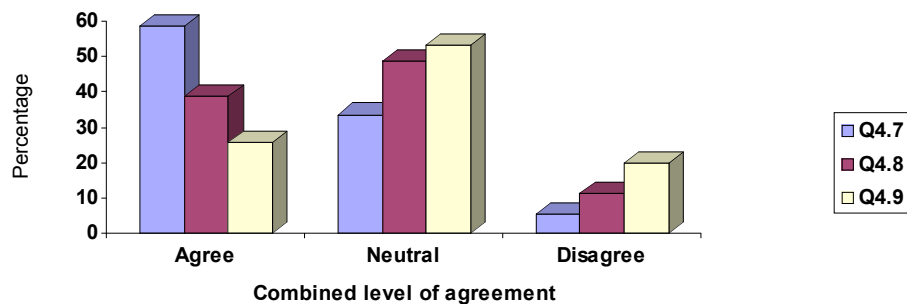
The Operation of the Peer Review Panel Section Summary

901 Generally, a very positive response was obtained for the following statements, with the overwhelming majority of respondents in agreement: 'Overall, the peer review panel process was sound' (92.9%), 'The composition of my peer review panel ensured that we made balanced judgements' (88.6%), 'The process by which my peer review panel allocated numerical scores was fair' (90.7%), and 'The process by which my peer review panel allocated interim grades was fair' (90.6%) was very positive with the majority of respondents in agreement. Of particular note was the response to the statement 'The TEC secretariat supported me and my panel effectively' where almost two thirds

(62.9%) of respondents strongly agreed and a further 35.0% agreed resulting in a combined total of 97.9%.

902 Higher levels of uncertainty were evident in the statements 'The moderation process ensured that the allocation of final grades was consistent across disciplines' (Q4.7), 'Participating TEOs will accept the final grades were allocated fairly' (Q4.8) and 'Participating researchers will accept the final grades were allocated fairly' (Q4.9). As the Figure 6 below indicates, relatively high numbers of respondents neither agreed nor disagreed. This uncertainty was particularly evident in responses to Q4.9 regarding individual researcher's acceptance of the final grades¹⁰.

Figure 6: A Comparison of Agreement with Statements about the Fair Allocation of Grades



903 In very general terms, the highest percentage of members agreeing with the statements in this section came from the HEA panel whereas the lowest percentage in agreement tended to come from the BEC and MED panels.

904 Only in the responses to question 4.3 (The panel member's Deskfile was a useful tool in preparing me for my role) did the percentage of non-chairs (72.7%) in agreement outweigh chairs (66.7%). **In all other cases, a higher percentage of chairs agreed.**

¹⁰ See the Qualitative Results, Part 2 for further discussion.

Section 5: Personal Experiences of being on the Panel.

905 This final section sought information regarding the respondents' personal experiences of being on the peer review panels. This included issues surrounding workload and the various forms of support they received.

Q5.1 My panel worked well together.

906 Almost all the respondents indicated some level of agreement with the statement 'My panel worked well together' with almost one third (32.1%, n=45) agreeing and almost two thirds (65.7%, n=92) strongly agreeing. Only 2.1% (n=3) were neutral and no respondents disagreed.

Table 30: Frequency table for Q 5.1 - My panel worked well together.

	Frequency	Percent	Cumulative Percent
Strongly agree	92	65.7	65.7
Agree	45	32.1	97.9
Neither	3	2.1	100.0
St.Agree/Agree	137	97.9	
Total	140	100.0	

907 All members of most panel agreed that their panel worked well together with the exceptions of BEC (at 92.3%) and SOC (at 86.7%).

908 Cross-tabulations showed all 100% of chairs agreed/strongly agreed with this statement, as did a similar proportion (97.7%) of non-chairs.

Q5.2 My workload as a panel member was reasonable.

909 Most of the respondents either agreed (58.6%, n=82) or strongly agreed (24.3%, n=34) that their workload as a panel member was reasonable, however, there were some who disagreed (10.7%, n=15) or strongly disagreed (2.1%, n=3).

Table 31: Frequency table for Q 5.2 - My workload as a panel member was reasonable.

	Frequency	Percent	Cumulative Percent
Strongly agree	34	24.3	24.5
Agree	82	58.6	83.5
Neither	5	3.6	87.1
Disagree	15	10.7	97.8
Strongly Disagree	3	2.1	99.3
St.Agree/Agree	116	83.5	
St.Disagree/Disagree	18	12.8	
Missing	1	0.7	
Total	140	100.0	

910 All those from the HEA and MIS panels agreed with the statement 'My workload as a panel member was reasonable', however, only 33.3% of those from CPA agreed.

911 Cross-tabulations showed all 100% of chairs agreed/strongly agreed with this statement compared with a slightly lower figure (82.0%) of non-chairs.

Q5.3 I was well prepared.

912 **Almost all** the respondents either **agreed** (56.4%, n=79) or **strongly agreed** (34.3%, n=48) with the statement 'I was well prepared'. This resulted in a combined total of 90.7% (n=127) with most of the remaining respondents neither agreeing nor disagreeing (6.4%, n=9).

Table 32: Frequency table for Q 5.3 - I was well prepared.

	Frequency	Percent	Cumulative Percent
Strongly agree	48	34.3	34.3
Agree	79	56.4	90.7
Neither	9	6.4	97.1
Disagree	3	2.1	99.3
Strongly Disagree	1	0.7	100.0
St.Agree/Agree	127	90.7	
St.Disagree/Disagree	4	2.8	
Total	140	100.0	



- 913 All members of the BS, EDU, HAL, HEA and MIS panels agreed that they were well prepared. This compared with only 50.0% of those from CPA.
- 914 Cross-tabulations showed all 100% of chairs agreed/strongly agreed with this statement compared with a slightly lower proportion (89.8%) of non-chairs.

Q5.4 I understood my role as a panel member.

- 915 A **relatively high** combined total of 96.4% of the respondents either agreed (51.4%, n=72) or strongly agreed (45.0%, n=63) with the statement 'I understood my role as a panel member'. The remainder neither agreed nor disagreed (3.6%, n=5).

Table 33: Frequency table for Q 5.4 - I understood my role as a panel member.

	Frequency	Percent	Cumulative Percent
Strongly agree	63	45.0	45.0
Agree	72	51.4	96.4
Neither	5	3.6	100.0
St.Agree/Agree	135	96.4	
Total	140	100.0	

- 916 All members of most panels agreed with the statement 'I understood my role as a panel member'. The exceptions were BEC, MED, MKD and SOC.
- 917 Cross-tabulations showed all 100% of chairs agreed/strongly agreed with this statement, as did a similar proportion (96.1%) of non-chairs.

Q5.5 I received any advice or help that I required.

- 918 Almost the same number of respondents either agreed (48.6%, n=68) or strongly agreed (47.1%, n=66) that they received any advice or help that they needed. Only 2.9% (n=4) were neutral and only 1.4% (n=2) disagreed.



Table 34: Frequency table for Q 5.5 - I received any advice or help that I required.

	Frequency	Percent	Cumulative Percent
Strongly agree	66	47.1	47.1
Agree	68	48.6	95.7
Neither	4	2.9	98.6
Disagree	2	1.4	100.0
St.Agree/Agree	134	95.7	
Total	140	100.0	

919 All members of the BS, CPA, ETA, HAL, MIS, MKD and PHY panels agreed that they received any advice or help they needed. A lower percentage from EDU agreed with this statement.

920 Cross-tabulations showed all 100% of chairs agreed/strongly agreed with this statement, as did a similar proportion (95.3%) of non-chairs.

Q5.6 I found being on the peer review panel a valuable professional experience.

921 The majority of the respondents (49.3%, n=69) strongly agreed that they found being on the peer review panel a valuable professional experience, and a further 42.1% (n=59) agreed. Of the remainder, 6.4% (n=9) were neutral with only 1.4% (n=2) disagreeing or strongly disagreeing (0.7%, n=1).

Table 35: Frequency table for Q 5.6 - I found being on the peer review panel a valuable professional experience.

	Frequency	Percent	Cumulative Percent
Strongly agree	69	49.3	49.3
Agree	59	42.1	91.4
Neither	9	6.4	97.9
Disagree	2	1.4	99.3
Strongly Disagree	1	0.7	100.0
St.Agree/Agree	128	91.4	
St.Disagree/Disagree	3	2.1	
Total	140	100.0	

922 All members of the BS, EDU, HEA, MKD and SOC panels agreed with the statement 'I found being on the peer review panel a valuable



professional experience', however, only 72.2% of those from the PHY panel did so.

- 923 Cross-tabulations showed all 100% of chairs agreed/strongly agreed with this statement, as did a similar proportion (90.6%) of non-chairs.

Personal Experience of Peer Review Panel Section Summary

- 924 Overall, the responses to the statements in this section were extremely positive. The combined total of those in agreement for each statement (with the exception of Q5.2) ranged from 90.7% (Q5.3) to 97.9% (Q5.1). Q5.2 'My workload as a panel member was reasonable' had a slightly lower level of agreement at 83.5%. Furthermore, there are indications in the qualitative data that although the workload per se may have been reasonable, the time frame in which the work was to be accomplished was too short¹¹. In all cases, a higher percentage of chairs agreed with the statements made in this section than non-chairs.

Part B: Responses to the Open-ended (Qualitative) Questions in the PBRF Evaluation On-line Survey

- 925 The final section of the PBRF Evaluation On-line Survey asked respondents to comment on aspects of their work on the panels, and the impact or outcomes of the PBRF that interested or concerned them. While there were several comments regarding points of interest, most of the data generated by the two open-ended questions related to aspects about which respondents were concerned or unsure. In the face of so many concerns and uncertainty it does pay to remember that, in the context of the quantitative results, most respondents were fairly positive about the PBRF model.
- 926 During analysis, each of the two questions was treated separately with the data categorised according to the various themes that emerged. While there was a degree of overlap in the answers to the questions, they are presented here as discrete entities.

Q1. Please comment on aspects of your work on the panels that interest or concern you.

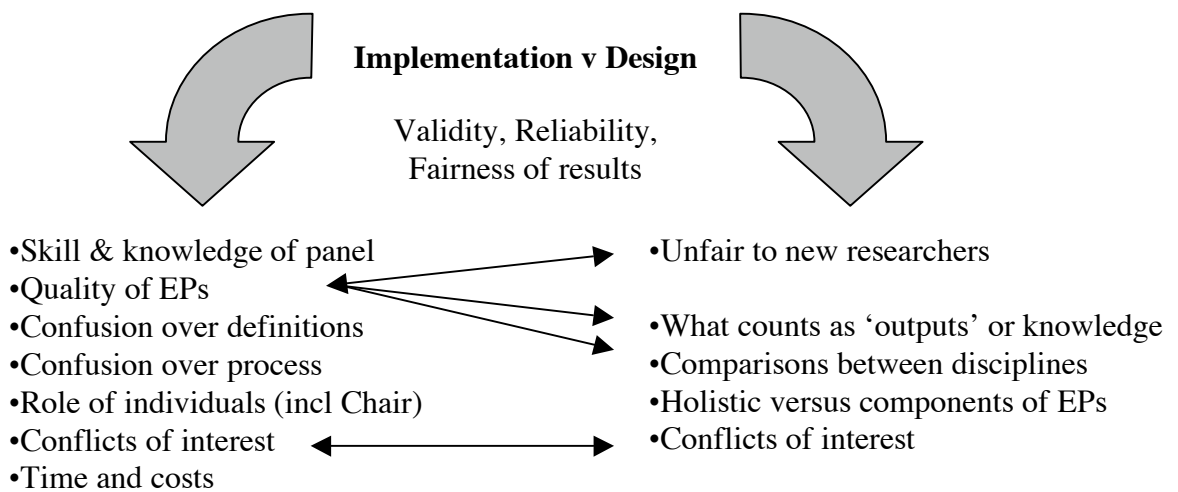
- 927 In total, seventy respondents made some comment in response to this invitation with many making several remarks, some of which were quite lengthy. Several observations, while brief, were clearly articulated with great feeling, thus the data gathered here was disparate in both content and intensity.

¹¹ See the Qualitative Results, Part 2.

928 Initially, the data were categorised very broadly in terms of whether the comment (or part thereof) related to the implementation of the PBRF or its design. This was seen as important in identifying early indicators of the success or otherwise of the model.

929 Once the data were assigned to one of these basic categories, sub-themes were then identified. These are depicted in Figure 7 and described in greater detail below. Importantly, these themes represent various ways in which the validity, reliability and fairness of the results of the PBRF Quality Evaluation might be compromised.

Figure 7: Aspects of Implementation and Design that affect Validity, Reliability and Fairness of Results.



Implementation

1. Skills/knowledge

930 While a number of respondents pointed out that their panel had worked well together and managed to make balanced judgements a more typical comment was that, in some circumstances, the panel as a whole did not possess sufficient expertise in some disciplinary areas and was therefore over-reliant on certain panel members. It was also pointed out that panel members should have an understanding of the New Zealand funding situation.

2. Quality of EPs

931 The most frequent observation in this regard pertains to the variation in the quality, standard or content of the Evidence Portfolios submitted by the participating TEOs. This was sometimes construed as deliberate 'massaging' on their part, but in many cases it was attributed to a lack of attention to the basic requirements or misunderstandings of definitions. . Such variation made valid comparisons difficult.

932 It was also noted that this round of the PBRF was more a reflection of the quality of the EPs than a measure of the quality of research

3. Confusion over definitions and guidelines

933 Another factor that can interfere with a valid and reliable result was the degree of confusion over either definitions or applications of particular terms. The terms 'quality assured' and 'world-class research' were mentioned as problematic on several occasions as was the rather important distinction between PE and CRE. Several participants commented that they were confused about what should be given 'special consideration', but others lamented that it could not be applied more widely, particularly in the face of heavy teaching loads. Finally, concern was also expressed over the right of panel members to argue that a paper be downgraded based on their personal opinion despite its acceptance by reviewers and editors.

4. Role of individuals

934 The most important point made here related to the influence of some panel members or chairs. This influence could be experienced in either positive or negative terms. Some respondents complimented their chair for doing 'an excellent job' whereas others were accused of impacting upon the evaluations in inappropriate ways. One respondent observed

that 'Individual panel members influenced evaluations within their discipline unduly'.

5. Conflict of interest (also a *design* issue)

935 The issues surrounding conflicts of interest can be described in terms of implementation and of design. The problem was typically attributed to the small population of New Zealand academics and the inevitability of having to play a role in evaluating people one knew or even worked with. The following quotations typify particular concerns:

- It is difficult to assess a colleague in the same Faculty [and]...it is almost impossible for the assessor to demonstrate that there is no bias.
- [Evaluating other panel members' EPs] was very awkward and it was clear that people felt reluctant to argue against upgrading others in the group in case the same arguments were used against them.
- I felt panel members were relying on their personal knowledge of a reviewee's work rather than strictly on what was in the EP.

6. Time/cost

936 While the workload placed on panel members did not seem to incur much comment, the short timeframe was a recurring theme. This is significant because as one respondent neatly summarized, 'The highly compressed timescale for the EP evaluations could have placed the quality of the process at moderate risk'.

Design

- 937 While the above points regarding the implementation of the PBRF can possibly be remedied by further experience with the new system and some modifications, numerous comments were also made concerning the design or assumptions underlying the model that might lead to a more systematic failure of the PBRF to effectively evaluate the quality of research conducted in New Zealand.

Unfair to new researchers

- 938 Perhaps the most common remark made in response to this open-ended question was directed at the Quality Categories and their relevance for new researchers. As one respondent describes 'New researchers in the first few years of their careers are certainly disadvantaged in terms of PE and CRE compared with more established researchers'.
- 939 Such observations characteristically led to some rather harsh criticism of the 'R' grade in particular with many advocating, for example, an 'N' rating for new researchers.

What counts as 'outputs' / knowledge

- 940 Although comments relating to what counts as knowledge were less common than comments on other issues, they are interesting because of the potential impact on the New Zealand research environment of this question. There were a number of issues involved.
- 941 First, it was noted that 'quality' research is not confined solely to specific journals and that 'excellent work does sometimes pop up in unexpected quarters'. There was concern that some research of a very high standard may have been overlooked.
- 942 Second, the evaluation of original creative work and Maori/Pacific Island material presented problems for some evaluators. Work that was of great significance at a national or regional level was sometimes seen as being undervalued because it was not regarded as 'world class'.
- 943 Third, a number of comments were made about 'valuable applied research being downgraded relative to pure research published in overseas journals'.
- 944 Fourth, cross-disciplinary comparisons could be difficult especially when proxy measures of quality were used that may, in fact, be specific to particular disciplines.

945 Fifth, it was noted that emergent disciplines, or work of an inter-disciplinary nature, might be disadvantaged (and therefore discouraged) by a lack of established benchmarks of quality.

946 Finally, a further disincentive for new researchers was the perceived rule that a PhD thesis was not considered evidence of research. Related to this point was the observation that 'having post docs will be a PBRF burden' despite their often valuable contributions.

Comparisons between disciplines

947 Although there were numerous comments about the success of evaluating and comparing research quality within disciplines, sometimes conflicting accounts were given of the feasibility of trying to do this across disciplines. While some urged a rather cautious treatment of the outcomes of the evaluation, others appeared confident that their results represented valid measures of research quality across a range of disciplines.

Holistic versus components of EPs

948 Several respondents indicated they were uncertain about the possibility of approaching evaluation decisions in a more holistic fashion. As one respondent noted 'The scoring system meant that PE and CRE was often the difference between getting an A grade or not. Some element of overall holistic assessment was required to make this judgement part of the scoring system. This should be made explicit in the process'.

Conflict of interest (see above)

Q2. Please comment on the outcomes or likely impacts of the quality evaluation that interest or concern you.

949 As in Q1 above, this invitation generated a wide range of comments and concerns from the respondents. Unlike the first question, however, these comments did not appear to be easily separated into issues of design versus implementation and were considered to be issues of design. The comments were categorised as follows:

- Alternatives
- Grades
- Funding
- Morale

- Individuals/Institutions
- Quality/Teaching
- Research Areas
- Competition

Alternatives

950 The respondents offered a substantial number of alternatives to the PBRF Quality Evaluation Framework. While they displayed considerable variation, the suggestions generally pertained to either what was assessed (e.g. citations which would be 'cheaper and simpler'), the unit of assessment (the individual versus the department, for example), the grade system (particularly a new 'less pejorative' term for new researchers) or the weightings given to each of the three quality indicators. With regards to this last point, although there were various suggestions it cannot be said that there was consensus over how the weightings should be distributed. Some opted for greater importance to be placed on CRE while others argued for less.

Funding

951 Although the remarks about funding were quite diverse, one theme to emerge was the lack of any guarantee that funding would necessarily end up in the hands of those who earned or deserved it. It was commonly noted that there was a lack of clarity over how PBRF funds were to be used and that, unless this became clear, the funds might not serve as any real incentive to improve research quality. Other comments related to the idea that using individual portfolios to arrive at an institutional formula was not completely logical while others were not entirely convinced that concentrating research funding in existing areas of research would ultimately raise the average level of research.

Quality Categories

952 There was some concern that the Quality Categories were too broad or, as one respondent put it, 'The difference on the boundary between A&B or B&C grades is almost insignificant yet a line is drawn and a negligible point of difference translates into a whole grade of difference'. Another stated, 'Clearly within the B grade there is a wide discrepancy between someone who has nearly made an A and someone who is solid B'. Under these circumstances it might be more useful for people to know their scores so that they can gain a sense of where they are within those categories and where to apply their efforts in order to raise their grade.

Some explained that a finer grading system might prevent people from becoming unhappy or demoralised when presented with their grade.

- 953 Others expressed a deeper skepticism with the process, a skepticism typified in the comment that 'The scoring and ultimately the grading into A B and C grades is a particularly ill thought out way of providing public information about the REAL research level attained in NZ'.

Morale

- 954 The likely effects of the PBRF on morale are neatly summarised by this account:

The effect of the assessment exercise on my colleagues is clearly negative. Those that are As will have their status confirmed those that are Bs will generally be irritated that they are not As. Cs, particularly young staff, will be very demotivated as they will see it as their efforts not being appreciated. The effects on the Rs can only be imagined. More important is the universal feeling that this is a radical and very undesirable ranking of academics.

- 955 This respondent was certainly not alone in predicting a significant negative effect on new researchers who will get a 'very negative message of how their efforts are valued'.

Individuals/Institutions

- 956 There was a degree of conflict in the various predictions of the impacts of the quality evaluation on both individuals within institutions and on the institutions themselves. Some believed that the evaluation process had led to a fair assessment of the quality of an individual's research and should therefore be used for recruitment purposes and for encouraging individuals to 'own' their grades and become responsible for improving it.

- 957 Others argued that there were compelling reasons for not reporting individual scores to individuals and TEOs. The reasons given for why they should not be used included the degree of subjectivity that was involved in arriving at the scores which could lead to flawed results, and the prospect of 'talent-buying' which was described as a 'lamentable practice'. One respondent doubted that the Quality Categories would remain secret in such a small country and that there would be considerable impacts on the recruitment and retention of staff. Another concern was that individuals would concentrate on 'grade chasing' rather than teaching. This was at the core of many fears that the PBRF will encourage individual rather than collaborative research efforts as academics compete to produce more research.

Quality

- 958 It was interesting to note the various ways in which these respondents believed the PBRF model would impact upon quality in research and teaching and whether competition would play a role. Each of these aspects is discussed in turn.

Teaching

- 959 There were numerous comments regarding the impact the PBRF would have on teaching as well as the impact teaching load would place on an individual's ability to produce high quality research. The following comment is characteristic of those who were nervous about the former:

I am concerned that staff will attempt to use their individual grades as leverage to buy out of their teaching and/or administrative responsibilities in a resource unit...and/or lobby for the differential allocation of funds in their resource unit to support their research.

- 960 Many expressed concern about a growing distinction between teaching and research staff where 'teaching will end up in the hands of part-time casuals whose cheapness makes it possible to create well-paid research-only positions'. On the other hand, there were also some concerns that strong teachers would be penalised by the system due to an increased pressure to do more research.

Areas of research

- 961 While concerns about the impact of the PBRF on particular areas of research were not as widespread as, for example, disquiet about the impact of R ratings for new researchers, those who did comment were seriously worried about the possible effects on particular types of research. The general concern was that the PBRF inadvertently discouraged or even penalised interdisciplinary or nascent research areas.
- 962 This was summarised by one respondent who noted 'chasing old money with new confirms existing research paradigms - it does not encourage new objects or methods of research' because it is more difficult to establish 'hard criteria' for evaluating quality in these areas. Research directed at areas specifically relevant to New Zealand were also seen as particularly vulnerable and may end up 'struggling to survive, despite their strategic importance, and could well find themselves in a downward spiral from which there is no escape'. Many of these concerns are described in this observation:

In established fields there are more established scholars resulting in a higher rating for the academic unit. Further, the quality of the research outlets, the markers of PE etc., are clearly established, widely known and agreed upon. Disadvantaging new research areas might encourage stagnation and discourage innovation upholding traditional Mandarin scholarly hierarchies. Students might be deprived of opportunities to explore new fertile areas of scholarship and research because universities would be reluctant to develop new programmes. Long-term research would suffer because new areas would not be explored and deemed too risky. I am also concerned that the quality evaluation might encourage universities to disestablish service areas that support research and develop student knowledge but that are not necessarily very research active. An example would be courses that were clinically oriented. Because quality is difficult to establish and quantity is incontrovertible I am concerned that scholars will feel compelled to publish as much as possible as quickly as possible.

Competition

- 963 There were several sub-categories within the overriding theme of competition which varied according to whether one was looking at the individual, institutional or global level. Although there was a degree of consensus that the PBRF process will 'encourage competition between, rather than collaboration among, different researchers' opinion was more divided as to whether this would apply to co-operation or competition between institutions.
- 964 One respondent noted that the PBRF 'does nothing whatever to help collaborative or inter-institutional research which is essential if we wish to remain or even gain international status'. Others were more optimistic about the PBRF in terms of 'transforming universities into being more competitive internationally' and being more 'strategic...with an increased focus on prioritisation of research themes and team building'. Another noted that 'The process is a highly competitive one and in the short term may change research cultures within institutions to meet PBRF methodological and quality evaluation requirements rather than build capacity and capability'. It is too soon to predict how these themes will be played out in future.

Conclusion

Is the design of the PBRF appropriate?

965 The PBRF appeared to satisfy the majority of respondents with over 90% agreeing that the PBRF framework allows for valid assessments of research quality to be made in both principle and in practice. Furthermore, nearly three quarters agreed that the assessment of individual EPs would encourage individual research performance, at least in principle. We can conclude, therefore, that in a very general way, the PBRF framework appears to be appropriate in the eyes of this particular population of panellists. We cannot extrapolate such opinions to the rest of the academic world, however.

Are the indicators of research quality valid and reliable?

966 Although the majority agreed that the three components of the Evidence Portfolios (NROs, PE and CRE) did constitute valid indications of research quality, there was less consensus that the weighting between them was right. The qualitative data suggest varying opinions and reasons as to which of the three should be given more or less weighting.

967 Related to this point were concerns about the PBRF model's ability to provide valid comparisons of research quality and achieve consistency in its processes *across* different disciplines.

Are the letter grades adequate discriminators of research quality?

968 Lower numbers of respondents agreed that the letter grades (A, B, C and R) provided adequate differentiation of the range of research quality and the qualitative results suggest a number of problems. The categories are so broad that the individual researcher cannot adequately gauge their position and, further, for those on the margins the distinction between two grades seemed minimal.

969 The 'R' grade was also seen as problematic, and many believed it to be a demoralising result for new researchers in particular. The results suggest that, for these respondents, finer gradations of research quality might provide better discriminators of research quality.

Was the assessment of EPs conducted in a robust, consistent and fair manner?

970 The majority (in excess of 88%) agreed that the peer review panel process was sound, that the composition of the panel ensured balanced judgements were made, and that the process of allocating numerical

scores and interim grades was fair. This suggests a reasonable level of faith that the results were indeed fair. However, the qualitative data did reveal some concerns about possible conflicts of interest and how these might undermine the overall even-handedness of results.

Other issues

- 971 There were a number of issues raised in comments provided in the open-ended questions section besides conflict of interest and the impact of R grades on new researchers. These could be divided according to whether the comment related to the design of the PBRF, its implementation or impact. These included the level of skills or knowledge of the panel, the disparate quality of the EPs which made comparisons difficult, confusion over definitions and guidelines, the short timelines which could compromise fair processes, the futility of trying to compare the outputs of different disciplines and what is counted as research.
- 972 This last issue was seen as having a potentially negative effect on new areas of research, interdisciplinary research, New Zealand/Pacific Island research and the creative and performing arts. The PBRF was seen as discouraging research in these areas because there is a lack of established criteria of what constitutes 'quality'. There was also concern directed at the impact the PBRF might have on teaching and on fostering a distinction between 'cheap' teachers and well-paid research-only positions.



On-line Survey form

Welcome to the on-line survey being undertaken as part of the evaluation of the PBRF. Your comments will make a crucial contribution to the evaluation, particularly given your experience as a Panel Chair or Member.

Every effort has been made to make sure the survey is carried out confidentially, and that individual responses cannot be uniquely identified.

The questionnaire will take approximately 10 minutes to complete. The questions in the first five sections ask you to click on the button that best represents your answer. At the end of the questionnaire there are two open-ended questions that give you the opportunity to expand on any of your ticked responses, or raise other issues with the WEB Research Evaluation Team.

When you have completed the questionnaire, please click the 'Done' key at the bottom. Until 'Done' is pressed, your answers will not be saved. Once you have clicked 'Done', the questionnaire is completed and your WWW browser will take you to the homepage of WEB Research.

Thank you for your time.

For sections 1-5, please indicate whether you agree or disagree with the following statements by clicking on the button that best fits your answer.

Strongly Agree Agree Neutral Disagree Strongly Disagree

1. The first set of statements relates to the Performance Based Research Fund Model.

- 1.1 The design of the PBRF quality evaluation framework, in principle, allows for valid assessments of research quality to be made.
- 1.2 Based on my experience of the implementation of the PBRF quality evaluation framework, in practice, valid assessments of research quality can be made.
- 1.3 The PBRF model is, in principle, a sound funding mechanism.
- 1.4 Assessing the Evidence Portfolios (EPs) of individuals, in principle, will encourage improved research performance.

2. This set of statements relates to indicators of research quality and the Quality Categories.

- 2.1 Nominated Research Outputs (NROs) are valid indicators of research quality.



- 2.2 Peer Esteem (PE) is a valid indicator of research quality.
- 2.3 Contribution to the Research Environment (CRE) is a valid indicator of research quality.
- 2.4 The combination of NROs, PE and CRE is a valid indicator of research quality.
- 2.5 The weighting between NROs, PE and CRE (70:15:15) is about right.
- 2.6 It is appropriate to translate scores into grades.
- 2.7 Using the letter grades (A B C R) resulted in an adequate differentiation of the range of research quality in the EPs submitted.
- 3. This set of statements relates to the impact of the quality evaluation of individual's Evidence portfolios.**
 - 3.1 Evidence Portfolios (EPs) provide a valid measure of individual research performance.
 - 3.2 EPs allow valid comparisons of individual research performance to be made within the same discipline.
 - 3.3 EPs allow valid comparisons of individual research performance to be made across different disciplines.
 - 3.4 Processes associated with the quality evaluation of EPs were consistent within the same discipline.
 - 3.5 Processes associated with the quality evaluation of EPs were consistent across different disciplines.
 - 3.6 The individual results of the quality evaluation are unlikely to be used in making decisions about the terms or conditions of employment of researchers.
 - 3.7 The PBRF is likely to promote more individual research efforts than prior to the PBRF.
 - 3.8 The PBRF is likely to promote more collaborative research efforts than prior to the PBRF.
 - 3.9 The PBRF is likely to ensure that research funding is concentrated where research is of the highest quality.



4. This set of statements relates to the operation of your peer review panel.

- 4.1 Overall, the peer review panel process was sound.
- 4.2 The composition of my peer review panel ensured that we made balanced judgements.
- 4.3 The Panel Member's Deskfile was a useful tool in preparing me for my role.
- 4.4 The TEC secretariat supported me and my panel effectively.
- 4.5 The process by which my peer review panel allocated numerical scores was fair.
- 4.6 The process by which my peer review panel allocated interim grades was fair.
- 4.7 The moderation process ensured that the allocation of final grades was consistent across disciplines.
- 4.8 Participating TEOs will accept that the final grades were allocated fairly.
- 4.9 Participating researchers will accept that the final grades were allocated fairly.

5. This set of statements relates to your personal experience of being on the peer review panel.

- 5.1 My panel worked well together.
- 5.2 My workload as a panel member was reasonable.
- 5.3 I was well prepared.
- 5.4 I understood my role as a panel member.
- 5.5 I received any advice or help that I required.
- 5.6 I found being on the peer review panel a valuable professional experience.

6. Please comment on aspects of your work on the panels that interest or concern you.

7. Please comment on the outcomes, or likely impacts, of the quality evaluation that interest or concern you.



Appendix 3: Abbreviations

Abbreviation	Meaning
CRE	contribution to research environment
EFTS	equivalent full-time student
EP	evidence portfolio
ERI	external research income
FTE	full-time-equivalent
National Library	National Library of New Zealand
MOE	Ministry of Education
NAU	nominated academic unit
NRO	nominated research output
OAG	Office of the Controller and Auditor-General
PBRF	Performance-Based Research Fund
PBRF Census	PBRF Census: Staffing Return
PE	peer esteem
RAE	research assessment exercise
RO	research output
RDC	research degree completions
SDR	Single-Data Return
TEAC	Tertiary Education Advisory Commission
TEC	Tertiary Education Commission
TEO	tertiary education organisation



Appendix 4: Glossary of terms

Assessment period The period between 1 January 1997 and 31 December 2002. Only research outputs produced in this period are eligible for inclusion in Evidence Portfolios for the 2003 Quality Evaluation.

Census date 31 July 2003 (see PBRF Census: Staffing Return).

Contribution to research environment (CRE)

Contribution that an eligible staff member has made to the general furtherance of research in his/her TEO or in the broader sphere of their subject area. One of the three components of an Evidence Portfolio.

Eligible staff member

TEO staff member eligible to take part in the Quality Evaluation.

Evidence Portfolio Collection of information on an eligible staff member's research outputs, peer esteem, and contribution to the research environment during the assessment period; is reviewed by a peer review panel and assigned to a Quality Category.

External research income (ERI)

Income for research purposes gained by a TEO from external sources. External research income is one of the three elements in the PBRF funding formula, along with the Quality Evaluation and research degree completions.

Nominated academic unit

Groupings of staff as nominated by each TEO for the purposes of reporting aggregated results of the Quality Evaluation.

Nominated research outputs (NROs)

The (up to four) best research outputs that the eligible staff member nominates in her/his Evidence Portfolio. Given particular scrutiny during the Quality Evaluation process.



Other research outputs

The additional (up to 50) research outputs submitted by the eligible staff member as part of her/his Evidence Portfolio.

PBRF Census: Staffing Return

A process run by the Ministry of Education whereby TEOs provide a detailed census of those of their staff participating in the PBRF Quality Evaluation process.

Peer esteem (PE) Esteem with which an eligible staff member is viewed by fellow researchers. One of the three components of an Evidence Portfolio.

Peer review panel Group of experts who evaluate the quality of research as set out in individual Evidence Portfolios. There are 12 peer review panels each covering different subject areas.

Quality Category A rating of researcher excellence that eligible staff are assigned to following the Quality Evaluation process. There are four categories – "A", "B", "C", and "R". Category "A" signifies researcher excellence at the highest level, and Category "R" represents research activity or quality at a level which is insufficient for recognition by the PBRF.

Quality Evaluation The component of the PBRF that assesses the quality of research output produced by eligible staff, the esteem with which they are regarded for their research activity, and their contribution to the research environment.

Quality score A standard measure of research quality. It is calculated by adding the weighted Quality Categories (ie "A" [10], "B" [6], "C" [2], and "R" [0]) of the PBRF-eligible staff in a particular unit (such as a TEO, nominated academic unit, or subject area) and dividing by the number of staff concerned, either on a head-count or FTE basis.

Research degree completions (RDC)

A measure of the number of research-based postgraduate degrees completed within a TEO where there is a research component of 0.75 EFTS or more. One of the three elements in the PBRF funding formula, along with the Quality Evaluation and external research income.



Research output (RO)

Product of research that is evaluated during the Quality Evaluation process. One of the three components of an Evidence Portfolio.

Specialist adviser Expert in a particular subject area used to assist a peer review panel to evaluate a particular Evidence Portfolio.

Subject area An area of research activity. For the purposes of the Quality Evaluation, research activity was classified into 41 subject areas each of which embodies a recognised academic discipline or disciplines. The 41 subject areas are listed in Appendix H of *The 2003 Assessment*.

Tie-points The quality standards expected for scores 2, 4 and 6 in each of the three components of an Evidence Portfolio.

Validation and Verification of Staff Eligibility

The process agreed between the MOE and TEC whereby the MOE undertook an audit of PBRF staff eligibility. See Appendix C of *The 2003 Assessment* for commentary on the four audits carried out by the TEC in 2003.



Appendix 5: Reports reviewed

This list which is not exhaustive, should be read in conjunction with the references for the report.

All relevant TEAC reports

Sector reports to the TEC

- Auckland University of Technology
- Christchurch College of Education
- Lincoln University
- Massey University
- Otago University
- Te Wānanga O Aotearoa
- University of Auckland
- University of Canterbury
- University of Waikato
- Victoria University

Workstream reports by the PBRF Team to the TEC

- Project Manager's Report – Project Closure Report
- PBRF Stakeholder Relationship Management – Workstream Closure Report
- PBRF Information Systems – Project Closure Report

Peer review panel reports to the TEC

- Report of the Moderation Panel
- Report of the Biological Sciences Panel
- Report of the Business and Economics Panel
- Report of the Creative and Performing Arts Panel
- Report of the Education Panel
- Report of the Engineering, Technology and Architecture Panel



- Report of the Health Panel
- Report of the Humanities and Law Panel
- Report of the Māori Knowledge and Development Panel
- Report of the Mathematical and Information Sciences and Technology Panel
- Report of the Medicine and Public Health Panel
- Report of the Physical Sciences Panel
- Report of the Social Sciences and Other Cultural/Social Studies Panel
- Moderation Panel – The 2003 Assessment
- First Moderation Panel Meeting: Issues Paper
- First Moderation Panel Meeting: Report on Audit Issues
- Second Moderation Panel Meeting: Issues Paper

Unpublished papers

Various unpublished papers (at the time of writing this report) presented to the Royal Society PBRF Forum, 21 May, 2004.



Appendix 6: Interviews, panel meetings and other events attended

Interviews during the scoping and in-depth study phases were held with

- PBRF Project Team Manager
- PBRF Operations and Panel Management Manager
- PBRF Sector Liaison Manager
- PBRF Systems Manager
- PBRF Policy and Design Manager
- Members of the Secretariat that supported the peer review panels
- Attendance at a weekly teleconference of TEO-based PBRF Project Managers

and staff in

- Ministry of Education
- Tertiary Education Commission
- Wellington College of Education
- University of Canterbury
- Lincoln University
- Association of University Staff
- Victoria University of Wellington
- Institutes of Technology and Polytechnics of New Zealand (ITPNZ)
- New Zealand Vice-Chancellors Committee - Research Sub-committee

This included TEO-based PBRF Project and Research Managers, as well as chairs and members of internal (TEO) panels or assessment processes.

Evaluation Advisory Group meetings attended

WEB Research attended meetings that were held approximately six weeks apart between the period October 2003 and June 2004. Draft material was also circulated for comment throughout that period.

Peer review panel meetings scheduled on the dates shown



See Appendix 1 for a breakdown of the actual sessions and numbers of researchers in attendance at each of the panel meetings.

Education Panel	18-21 November 2003
Humanities and Law Panel	24-27 November 2003
Māori Knowledge and Development Panel	18-21 December 2003
Medical and Public Health Panel	7 December 2003
Physical Sciences Panel*	19 December 2003
Moderation Panel	12 March 2003
	15 November 2003
	16 November 2003
	15 January 2003

Moderation panel meetings to recalibrate assessments by the Education, and Engineering and Technology and Architecture Panels, held in December 2003, and January 2004. * Afternoon session on date shown.

Case study sites

In the course of this evaluation, we visited six case study sites. We do not list them here in order to preserve their anonymity. The TEOs covered the following range:

- A large TEO
- A medium sized TEO
- A small TEO
- A university
- A college of education
- A wānanga

Interviews conducted in TEOs

We conducted 47 formal interviews in TEOs with staff who held a range of positions and qualifications. Positions included:

- PBRF manager or co-ordinator
- Financial/Administration/Information Technology support



- Research – Director, Manager, Support
- VC/VP Research
- Vice Chancellor, Principal, Vice Principal, Director, Dean
- Director/Chair/Head of Department or School or Division
- Senior Lecturer
- Lecturer
- Junior Lecturer
- Senior Teaching Fellow
- Research Fellow

Eleven of the participants either chaired or were panellists on their internal peer review panels.

Attendance at a Forum Evaluating the PBRF Assessment Framework hosted by the Royal Society of New Zealand Social Sciences Committee.

A forum on the design and implementation of the Performance Based Research Fund and its impact on New Zealand Universities held at the National Library of New Zealand on 21st May, 2004 hosted by the Royal Society of New Zealand Social Sciences Committee.