MINISTRY OF EDUCATION
Te Tāhuhu o te Mātauranga

Trends in measured research quality

An analysis of PBRF Quality Evaluation results

New Zealand Government

Research and knowledge creation

This report forms part of a series called Research and knowledge creation.

Author:
Warren Smart, Senior Research Analyst
Email: warren.smart@minedu.govt.nz
Telephone: 64-4-463-8035
Fax: 64-4-463-8713

Acknowledgements:
The author gratefully acknowledges comments provided by Roger Smyth (Ministry of Education), Jason Dowse (Ministry of Education), Nuran Çinlar (Tertiary Education Commission), Janet Grice (Tertiary Education Commission) and Professor Gary Hawke (Victoria University of Wellington) on earlier drafts of this report.

The author also gratefully acknowledges the work of Sean Alexander (Tertiary Education Commission) in constructing the database used in this report.

The author also thanks Virginia Falealili who proofread the report.

All views expressed in this report, and any remaining errors or omissions, remain the responsibility of the author.

Published by:
Tertiary Sector Performance Analysis and Reporting
Strategy and System Performance
Ministry of Education

This report is available from the Ministry of Education's Education Counts website:
http://www.educationcounts.govt.nz

July 2008

ISBN 978-0-478-13912-9 (Print)
ISBN 978-0-478-13913-6 (Web)

Trends in measured research quality

Contents

List of Figures and Tables

Glossary

Research output (RO) score
This is the score between 0 and 7 (with 7 representing the highest performance) awarded to staff by peer-review panels based on the quality of their research outputs.

Peer esteem (PE) score
This is the score between 0 and 7 (with 7 representing the highest performance) awarded to staff by peer-review panels based on the esteem in which they are held by their peers.

Contribution to the research environment (CRE) score
This is the score between 0 and 7 (with 7 representing the highest performance) awarded to staff by the peer-review panels based on their contribution to the research environment.

Weighted research score (WRS)
This is the weighted composite score of the RO, PE and CRE research scores. A weighting of 70/15/15 is applied to each of the scores, respectively, to arrive at the weighted research score. This score has a value between 0 and 700 (with 700 representing the highest performance). These scores are then used as a guide by the peer-review panels when assigning quality categories.

1 Summary

Key findings:

- This study analysed the change in measured research quality of around 2,000 staff who participated in both the 2003 and 2006 PBRF Quality Evaluations.
- Although all three average research component scores increased on average for these staff, the average peer esteem and contribution to the research environment scores increased much faster than the average research output score.
- The study also found that the prior performance of these staff in the 2003 Quality Evaluation was a key factor associated with improvement in their measured research quality, controlling for other factors. Staff with lower measured research quality in the 2003 Quality Evaluation, were more likely to improve their measured research quality in the 2006 Quality Evaluation.

This study used a mix of descriptive statistics and statistical modelling to identify how the demographic and employment-related characteristics of around 2,000 staff who participated in both the 2003 and 2006 Quality Evaluations were associated with changes in the level of their measured research quality.

Although other studies have also examined the change in measured research quality between the 2003 and 2006 Quality Evaluations,[1] this new study was more selective and focused solely on staff who had evidence portfolios assessed in both Quality Evaluations and who were not identified as new and emerging researchers. This group of staff were the only participants in the 2006 Quality Evaluation who had their research quality measured by peer-review panels in both Quality Evaluations and who were assessed under a similar scoring system. While this means that some important groups are omitted from the analysis, by focusing on this particular group of researchers, a clearer picture can be obtained of the components of change in measured research quality.

It should be noted that because staff and tertiary education organisations 'learned' from the experience of the first Quality Evaluation in 2003, the standard of presentation of the evidence portfolios improved in the 2006 Quality Evaluation and was a contributing factor to the improvement in measured research quality (Tertiary Education Commission, 2007). Because of this, part of any improvement in measured research quality identified in this study may be more a consequence of better presentation of evidence portfolios, than a 'real' lift in the quality of the research produced.

Two analytical approaches were used in this study to identify the factors associated with changes in measured research quality. Firstly, descriptive statistics were used to examine changes in measured research quality. The results of the descriptive statistics analysis showed that:

- Between 2003 and 2006, the average peer esteem (PE) and contribution to the research environment scores (CRE) increased by much more than the increase in average research output (RO) score. This greater improvement in the PE and CRE score, compared with the (RO) score, suggests that the improvement in measured performance was at least partly due to improved presentation in evidence portfolios. This is because there is a greater subjective element in the assessment of these dimensions. Nonetheless, there was also a rise in RO score, if to a lesser extent than the other research component scores. Given this research component is potentially less subject to change as a result of improved presentation of

[1] See Cinlar and Dowse (2008a, 2008b, 2008c) and White and Grice (2008).

evidence portfolios, this provides some evidence of an increase in quality of the research carried out by the staff selected in this study.

However, a drawback of using descriptive statistics to identify the factors associated with change in research performance is that, while it may appear that an association exists between a factor of interest and research performance, it may in fact be due to the influence of other confounding factors. To overcome this problem, this study then applied multiple regression to the sample dataset to further explore the association between the demographic and employment-related characteristics of staff and the change in measured research quality.

The results of the regression analysis showed that:

- The prior performance of staff in the 2003 Quality Evaluation was a key factor associated with the change in their measured research quality between 2003 and 2006. Staff with lower performance in the 2003 Quality Evaluation achieved a greater level of improvement in their research quality, holding other factors constant. Because those who scored higher in the 2003 Quality Evaluation were less likely to have a lift in their quality category, this suggests that there is a nonlinear relationship between measured research quality and the quality categories.

- Staff of higher academic rank achieved the greatest improvement in measured research quality, controlling for other factors.

- Older staff achieved smaller improvements in measured research quality than younger staff, controlling for other factors.

- Staff employed at the universities achieved greater improvements in measured research performance than staff at non-university tertiary education organisations, controlling for other factors.

- Staff who submitted a greater number of research outputs in the 2006 Quality Evaluation compared with the 2003 Quality Evaluation achieved greater improvement in measured research quality, controlling for other factors. However the scale of improvement from increasing the number of submitted research outputs was relatively small and was mostly restricted to staff of lower academic rank.

- In terms of subject area, staff in the 'Māori knowledge and development' panel had lower improvements in quality than other subject panels, controlling for other factors. However, the number of staff in this panel in the sample was small, so caution should be used in interpreting this result.

It is important to remember that this study analysed the performance of a select group of staff, those who participated in the 2003 and 2006 Quality Evaluations, had an evidence portfolio assessed both times and were not identified as new and emerging staff in 2006. Therefore, it should not be assumed that the findings of the analysis apply to all participants in the PBRF Quality Evaluations.

Finally, the way that the nature of the association between the increases in research quality within the various subgroups examined in this study changed following the application of regression analysis shows that care should be taken when using raw summary statistics to analyse any changes in performance. In particular, controlling for the level of measured research performance in 2003 was crucial to separating out the confounding effects of individual staff characteristics on measured research quality.

2 Introduction

The quality of research produced by staff in New Zealand's tertiary education organisations has been measured directly on two occasions. The first occasion was via the 2003 Performance-Based Research Fund (PBRF) Quality Evaluation and the second via the 2006 PBRF Quality Evaluation. This provides an opportunity to explore how measured research quality changed across a number of performance dimensions and across a variety of subgroups of interest.

The PBRF Quality Evaluations use a system of peer-assessment to measure the quality of research produced by staff at participating tertiary education organisations. The results of the PBRF Quality Evaluations are then used by the government to allocate funding to participating tertiary education organisations. In 2007, around $124 million[2] was distributed to tertiary education organisations based on their performance in the 2006 PBRF Quality Evaluation.

This study uses a mix of descriptive statistics and statistical modelling to identify the demographic and employment-related characteristics of participating staff associated with changes in the level of measured research quality between the 2003 and 2006 Quality Evaluations.[3] Although other studies have also examined the change in measured research quality between the 2003 and 2006 Quality Evaluations,[4] this new study is more selective and focuses solely on staff who had evidence portfolios assessed in both Quality Evaluations and were not identified as new and emerging researchers. This group of around 2,000 staff were the only participants in the 2006 Quality Evaluation who had their research quality measured by peer-review panels in both Quality Evaluations and who were assessed under a similar scoring system. The other participants either had their results from the 2003 Quality Evaluation carried over, had their evidence portfolios considered by a peer assessment panel in the Quality Evaluation for the first time, or were assessed as a new and emerging researcher in the 2006 Quality Evaluation.

While this means that some important groups are omitted from the analysis, by focusing on this particular group of researchers, a clearer picture can be obtained of the components of change in measured research quality. Given that this study focuses on only a quarter of the staff who participated in the 2006 Quality Evaluation, the results of this analysis should not be assumed to be representative of the other groups participating in the 2006 Quality Evaluation.

There are two analytical approaches used in this study to identify the factors associated with changes in measured research quality. Firstly, descriptive statistics capturing the average change in the measured research quality of staff between 2003 and 2006 are compared across four measures of performance: the percentage of staff improving their quality category, the average change in research output score, the average change in peer esteem score and the average change in contribution to the research environment score.

However, a significant drawback of using descriptive statistics to identify the factors associated with change in research performance is that, while it may appear that an association exists between a factor of interest and research performance, it may in fact be due to the influence of other confounding factors. To overcome this problem, this study applies multiple regression to the sample dataset to further explore the association between the demographic and employment-related characteristics of staff and the change in measured research quality.

The use of regression analysis allows us to examine the association between a characteristic of interest and the change in measured research quality while holding all other characteristics constant, helping to clarify the nature and degree of the association. In particular, staff who performed at a lower

[2] Of the $206 million of PBRF funding.
[3] Previous studies by Smart (2005; 2008) analysed the factors associated with research quality in the 2003 and 2006 PBRF Quality Evaluations.
[4] See Cinlar and Dowse (2008a, 2008b, 2008c) and White and Grice (2008).

level in the original 2003 Quality Evaluation, no matter what demographic or employment-related subgroup they belong to, would likely have more motivation (and room) to improve on their measured performance in the 2006 Quality Evaluation than staff who achieved higher levels of performance. In addition, tertiary education organisations may have directed more resources at lower-performing staff in order to improve their quality category and hence boost funding attracted via the PBRF. The regression analysis explores this issue by examining how much of the change in measured research quality is associated with the original level of performance by staff in the 2003 Quality Evaluation, rather than with the demographic and employment-related characteristics of staff.

It is important to note that because staff and tertiary education organisations 'learned' from the experience of the first Quality Evaluation in 2003, the standard of presentation of the evidence portfolios improved in the 2006 Quality Evaluation and was a contributing factor to the improvement in measured research quality exhibited by many staff (Tertiary Education Commission, 2007). Because of this, part of any improvement in measured research quality identified in this study may be more a consequence of better presentation of evidence portfolios, than 'real'. Nevertheless, it is still of interest to identify which groups of staff improved in measured research quality and across what dimensions of performance to get a sense of what was driving the changes in measured research quality.

The structure of this analysis is as follows. Firstly, a brief explanation of the process of allocating PBRF quality categories is presented in section 3. Then, the data and method used in this study is presented in section 4. An analysis of descriptive statistics is presented in section 5 and is followed by the presentation of the results of the regression analysis in section 6. Finally, some conclusions are presented in section 7.

3 Measuring research quality[5]

A similar process was followed in the 2003 and 2006 Quality Evaluations to determine the quality categories assigned to the sample of staff used in this study. In both Quality Evaluations, evidence portfolios were submitted to peer review panels. These portfolios provided evidence of the research performance of staff across three dimensions: the quality of their research outputs (RO), the esteem they are held in by their peers (PE) and their contribution to the research environment (CRE).[6]

After reviewing the evidence portfolios, the expert peer review panels allocated a score from 0 to 7 to each of the RO, PE and CRE dimensions after evaluating the content of the evidence portfolios, with 7 representing the highest performance and 0 the lowest. A 70/15/15 weighting was then applied to the three component scores and a weighted research score (WRS) between 0 and 700 calculated.

A WRS between 600 and 700 was generally assigned an A quality category, a WRS between 400 and 599 was assigned a 'B', a WRS between 200 and 399 was assigned a C, and a WRS between 0 and 199 an R.[7] In general terms, an A quality category indicated that a researcher was producing research that was esteemed internationally. A B quality category indicated that research was recognised at a national level for its quality and a C quality category indicated the quality of research was recognised at a local level within their tertiary education organisation. An R quality category indicates the staff have been assessed by the panel as not having met the requirements of a C quality category.

The sample of staff used in this analysis does not include staff who were identified as being new and emerging as the scoring system used to allocate a quality category to these staff changed in the 2006 Quality Evaluation. The different scoring system that was introduced for new and emerging researchers in 2006 is described below.

If a staff member was employed for the first time in an academic role after 1 Jan 2000 (either in NZ or overseas) or required to do degree teaching for the first time from this point, they could be nominated by their tertiary education organisation as a 'new and emerging' researcher. New and emerging staff were allocated quality categories using a slightly different approach. Like experienced researchers, new and emerging staff could be assigned an A or a B quality category. However, because new and emerging staff may not have had the opportunity to build up a significant portfolio of research, they could also be assigned a C(NE) or an R(NE) quality category. A C(NE) quality category did not require staff to submit evidence on the PE and CRE components, although they were encouraged to do so.

If new and emerging staff met certain reduced RO requirements they could still attract funding for their institution by being allocated a C(NE) quality category. It was expected that a minimum of two research outputs would be submitted in the evidence portfolio, in addition to a completed doctoral degree thesis.

Given the degree of difference in the scoring system used to assess new and emerging staff, a comparison of their scores between the two Quality Evaluations would be invalid.

[5] More detail on how quality categories were allocated can be found in Tertiary Education Commission (2007).

[6] Examples of what can be provided as evidence of PE include being awarded research-related prizes and invitations to present conference addresses. Examples of what can be provided as evidence of CRE includes the amount of external research income generated and the supervision of research students.

[7] As the peer-review panels followed a holistic process the quality category allocated could vary from the indicative WRS, although the number of cases where this happened was very small.

4 Data and method

4.1 Data

The dataset used in this report was supplied by the Tertiary Education Commission following an application to access unit record data from the Performance-Based Research Fund (PBRF) Quality Evaluations.[8] The dataset contained information on the quality scores and demographic and employment-related information of staff who participated in both the 2003 and 2006 Quality Evaluations.

The sample selected for this study contains information on the staff who submitted evidence portfolios in both the 2003 and 2006 Quality Evaluations and who were not identified as being new and emerging researchers. These are the only group of participants in the 2006 Quality Evaluation who had evidence portfolios assessed in both Quality Evaluations and who were assessed using a similar scoring system.[9]

This particular group was selected for this study as only those staff who had evidence portfolios assessed by the peer review panels in both Quality Evaluations received research component scores in 2003 and 2006 that could be compared. As was discussed in the previous section, even though some staff identified as new and emerging in the 2006 Quality Evaluation had received research component scores from the peer review panels in both Quality Evaluations, they were assessed under a different scoring system in 2006, making comparison of their performance between 2003 and 2006 invalid.

The 2006 Quality Evaluation was a partial round, with staff who had also participated in the 2003 Quality Evaluations not required to resubmit an evidence portfolio to the peer review panels. Their performance in the 2003 Quality Evaluation could simply be carried over. Therefore, these staff are excluded from this study, given that their research quality was not assessed in the 2006 Quality Evaluation.

A total of 2,025 staff were included in the sample used in this study. This is around 23 percent of the total number of staff who participated in the 2006 Quality Evaluation, 45 percent of staff who had evidence portfolios assessed in the 2006 Quality Evaluation and 75 percent of staff who submitted evidence portfolios in both 2003 and 2006.

Figure 1 presents the distribution of quality categories awarded in the 2006 Quality Evaluation for the sample of staff used in this study who resubmitted evidence portfolios and for all participating staff. Note that this analysis excludes new and emerging researchers, hence there are no C(NE) or R(NE) quality categories in Figure 1.

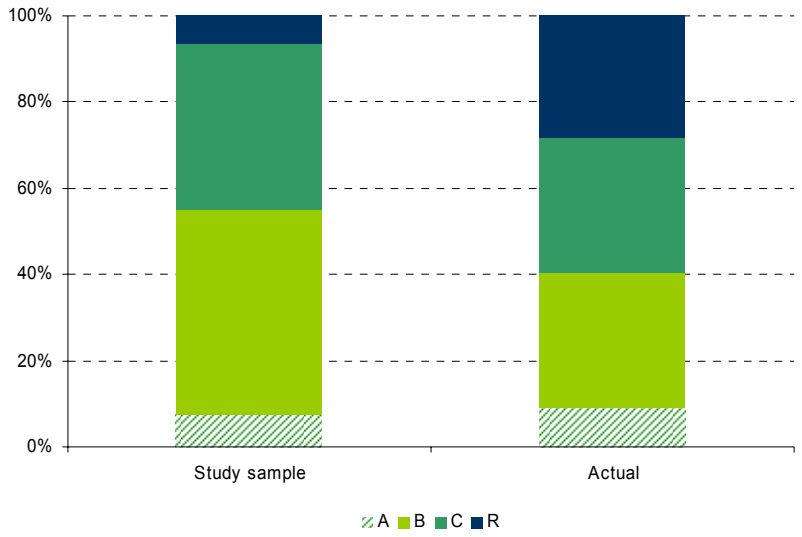
The sample used in this study contains a much smaller proportion of staff allocated an R quality category (6.7 percent) compared with the proportion of all participating staff who were allocated an R (28 percent). Evidence portfolios submitted to the peer review panels would generally have been considered by their tertiary education organisations to have a chance of receiving at least a C quality category. Hence the proportion of R quality categories in the sample used in this study is much smaller than among all staff who participated in the 2006 Quality Evaluation. Also, because almost all participating staff who received an A quality category in the 2003 Quality Evaluation had their quality

[8] See http://www.tec.govt.nz/upload/downloads/pbrf_dataaccess_final.pdf for more information on the process to access information from the PBRF Quality Evaluations.

[9] Excluded from the sample used in this study were two researchers that received an A quality category in the 2003 Quality Evaluation and who clearly could not improve on their 2003 result and for whom their motivation for resubmission would have been different to the other staff in the study. Also, because of missing values for age, 23 staff were excluded from the sample used in this analysis.

category carried over to the 2006 Quality Evaluation, there is a smaller proportion of As awarded in 2006 in the study sample.

Figure 1: Distribution of quality categories allocated in the 2006 Quality Evaluation (excluding new and emerging researchers)



4.2 Method

To identify the staff characteristics associated with changes in measured research quality two regression approaches are used. Firstly, binary logistic regression is used to analyse the association of staff characteristics with the likelihood of staff improving their quality category. Secondly, ordinary least squares regression is used to analyse the association of staff characteristics with the change in each of the research component scores allocated to staff.

4.2.1 Logistic regression

As a staff member's quality category either improved or not, the dependent variable in this analysis is binary in nature and the use of ordinary least squares is not appropriate.[10] Therefore, logistic regression, which uses a maximum likelihood procedure, is used to analyse the association between the demographic and employment related characteristics of staff and the likelihood of improving their quality category. If the staff member's quality category improved the dependent variable takes a value of 1, otherwise 0.

The logistic regression model is presented in equation 1. *HQC* represents the probability of a staff member achieving a higher quality category.

$$\ln\left[\frac{HQC}{(1-HQC))}\right] = \alpha \quad + X_i\beta + \mu \qquad (1)$$

Where the dependent variable is the natural logarithm of the odds of a staff member improving their quality category, X is a vector of explanatory variables, β is the coefficient of the explanatory variables in logit form, μ is an error term and i = 1 to n observations.

[10] Ordinary least squares regression is not appropriate where the dependent variable is binary in nature (Ramanathan, 1998).

The results of the logistic regression analysis are presented in the form of odds ratios. An odds ratio is the odds of an event happening divided by the odds of the opposite event happening. For example, suppose that 400 senior lecturers achieved a higher quality category and 200 did not. The odds of a senior lecturer achieving a higher quality category are 400/200 = 2, or 2 to 1. In other words, the chances of a senior lecturer achieving a higher quality category are reasonably good. To give another example, suppose that 500 lecturers achieved a higher quality category and 1,000 did not. The odds of a lecturer achieving a higher quality category would be 500/1,000 = 0.5, or 1 to 2. The chances of them achieving a higher quality category are therefore significantly lower than for senior lecturers.

For continuous explanatory variables, an odds ratio of greater than 1 indicates a higher likelihood of achieving a higher quality category as the value of the explanatory variable increases and a value less than 1 indicates a lower likelihood.

For categorical explanatory variables, the odds ratio compares the likelihood of achieving a higher quality category compared with the reference category. An odds ratio greater than 1 indicates a higher likelihood of achieving a higher quality category compared with the reference group, while a value of less than 1 indicates a lower likelihood.

The results of the logistic regression are also presented in the form of the change in predicted probability resulting from a change in the explanatory variable – the Delta P statistic. The Delta P statistic shows the percentage point change in the probability of staff achieving a higher quality category from a change in the explanatory variable, where the default predicted probability is set at the actual proportion of staff whose quality category improved (in this case 37.7 percent).

4.2.2 Ordinary least squares regression

Ordinary least squares regression is used to identify the variables associated with the change in the research component scores received by staff between the 2003 and 2006 Quality Evaluations. The dependent variable (*Q*) in this analysis is the change in the research component score allocated to staff members.

This can be modelled as:

$$Q_i = \alpha + X_i\beta + \mu \qquad (2)$$

where X represents the set of explanatory variables, β is the vector of regression coefficients, α is the intercept term, μ is an error term and i = 1 to n observations.

In this study there are three research component scores that are analysed, the research output (RO) score, the peer esteem (PE) score and the contribution to the research environment (CRE) score.

4.2.3 Explanatory variables

The demographic and employment-related characteristics used as explanatory variables in the regression analysis are discussed in turn below.

Previous research performance

The level of measured research quality achieved by staff in the 2003 Quality Evaluation is included in the regression model and is a key control variable. It is expected that staff who received relatively low levels of measured research quality in 2003 would have greater room and motivation to improve their performance compared with staff with higher levels of research quality. It may also be that staff with lower performance in the 2003 Quality Evaluation had more resources directed at them by tertiary education organisations attempting to increase the funding received via the PBRF. In addition, this

variable is included as it appears that, the higher the level of previous quality category, the harder it is for staff to move to a higher quality category.

For the logistic regression analysis, the quality category received by staff in 2003 is used to capture prior performance. The reference category is staff who received an R quality category in the 2003 Quality Evaluation. For the ordinary least squares regression the relevant research component score received in 2003 is used to capture prior performance.

Age

A variable that captures the age of staff at the time of the 2003 Quality Evaluation is included in the regression models. Staff who are nearing the end of their academic careers may be less motivated to improve their quality category than younger staff. Also, younger staff may be at a relatively new stage in their academic careers, so would be expected to improve at a faster rate than older staff who may have more experience in research. As a result, there may be a negative association between the age of staff in the sample and change in measured research quality.

Gender

In the regression models a dummy variable[11] is included to capture any gender-based effects. The reference group is male.

Ethnic group

A variable with multiple categories is included in the regression models to capture the ethnic group of staff. The categories are: European, Māori, Pasifika, Asian, Other and Unknown.[12] The reference group in the analysis is European. As staff could identify more than one ethnicity in the PBRF staff census the ethnic group data has been prioritised. The order of prioritisation is: Māori, Pasifika, Asian, Other and European. Note that because of the small number of Pasifika staff in the sample (18) caution should be used when examining the research performance of staff in this ethnic group.

Change in full-time equivalent status

Any change in full-time equivalent status between 2003 and 2006 is included as an explanatory variable in the regression models. It is expected, a priori, that staff who had an increase in their full-time equivalent status would improve on their measured research performance.

Position of staff

The position of staff at the time of the 2003 Quality Evaluation is included as a variable with multiple categories in the regression models. The categories are based on broad categories as defined by the Tertiary Education Commission and include: professor, associate professor, senior lecturer, lecturer, assistant lecturer and 'other' staff and administrative leader.[13] The reference category is senior lecturer.

Staff of higher academic rank who received a quality category similar to staff of lower academic rank may feel under more pressure to improve their quality category and therefore bring it into line with their academic status. Therefore, one might expect a higher likelihood of achieving a higher quality category to be associated with a higher academic rank, controlling for other factors.

Change in the number of submitted research outputs

The change in the number of research outputs submitted in evidence portfolios is included as a variable in the model. As staff could include up to 54 outputs in the 2003 Quality Evaluation, compared

[11] In this case if the staff member is a man the variable takes a value of 0 and if the staff member is a woman the variable takes a value of 1.

[12] Around 12 percent of staff in this sample had an unknown ethnic group. As a result of coding these staff as unknown, caution should be used when examining the difference in measured research quality between ethnic groups.

[13] For more detail on how these aggregated positions were determined see Çinlar and Dowse (2008a).

with 34 in the 2006 Quality Evaluation, those who submitted between 35 and 54 research outputs were recoded as having submitted 34 outputs.[14]

Type of institution
A variable that captures the type of institution the staff member was employed at is included in the regression model. There are two institution types in this variable: universities and other tertiary education organisations.

Although the Education Act 1989 states that the teaching of degrees is to be "…taught mainly by people engaged in research",[15] the characterisation of universities sets a higher threshold for research than for other tertiary education organisations. Universities are required to meet "international standards of research and teaching", whereas other education organisations are not.[16] Therefore, it is likely that staff employed at universities would have a greater motivation to improve the quality category they received compared with staff at other types of tertiary education organisations.

Given that staff may have moved between the two types of tertiary education organisations, the four categories in this variable are: 'Uni03/uni06', 'Other03/other06', 'uni03/other06' and 'Other03/uni06'. The reference category in this analysis is those staff who were employed at a university in both Quality Evaluations.

Subject
To capture the effect of subject area, a variable with multiple categories representing the broad PBRF subject panels in which a staff member was identified in 2003 was included in the regression models. The reference category is 'Biological sciences'. Note that because of the small number of staff in the sample in the 'Māori knowledge and development' panel (16), the results for this group of staff should be treated with caution.

The model was also re-estimated using a variable that captures the narrow subject area in 2003 so that the association between selected narrow subject areas and change in research quality can be examined.[17]

On the cusp of a higher quality category
This variable captures if a staff member was close to the cusp of achieving a higher quality category in the 2003 Quality Evaluation.[18] Staff who were on the cusp would have a higher likelihood of achieving a higher quality category in the 2006 Quality Evaluation. This variable takes a value of 1 if the weighted research score (WRS) of staff was within 50 points of achieving the next highest quality category, otherwise 0. The reference category was a staff member who was not on the cusp of a higher quality category.

This variable is only used in the logistic regression analysis as prior performance in the ordinary least squares regression analysis is captured by the research component score allocated in 2003. Therefore being on the cusp of a higher quality category is not relevant in that analysis.

[14] This variable is not weighted for the type of research output submitted in the evidence portfolio.
[15] Education Act 1989 section 254 (3a).
[16] Education Act 1989 section 162 (4.a.iii).
[17] The regression output from this re-estimation of the model using narrow subject areas is not reported but is available from the author on request.
[18] The issue of a staff member being 'on the cusp' was first discussed in Çinlar and Dowse (2008b).

5 Average change in measured research quality

This section gives summary statistics relating to the average change in measured research quality between the 2003 and 2006 Quality Evaluations by the staff included in this study. Note that a full table of summary statistics is provided in the Appendix at the end of this report.

5.1 Average change in measured research quality at the aggregated level

The distribution of the quality categories allocated to staff in 2003 and 2006 is illustrated in Figure 2 below. An improvement in the level of measured research quality is clearly evident with an increase in the proportion of staff awarded an A or B quality category in 2006 and a decrease in the proportion of staff awarded C and R quality categories. The proportion of staff allocated an R quality category decreased from 16 percent in 2003 to 6.7 percent in 2006. Similarly, the number of staff allocated a C quality category decreased from 47 percent in 2003 to 38 percent in 2006. The proportion of staff allocated a B quality category increased from 37 percent in 2003 to 47 percent in 2006. The proportion of staff allocated an A quality category in 2006 was 7.6 percent, remembering that the sample selected for this study did not contain any staff awarded an A quality category in 2003.

Figure 2: Distribution of quality categories

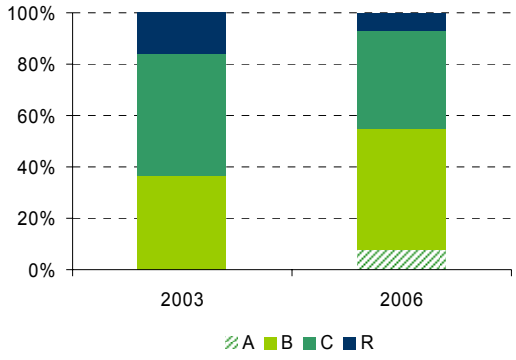
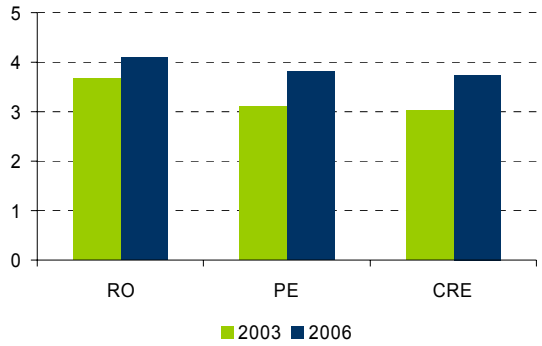


Figure 3: Average research component scores



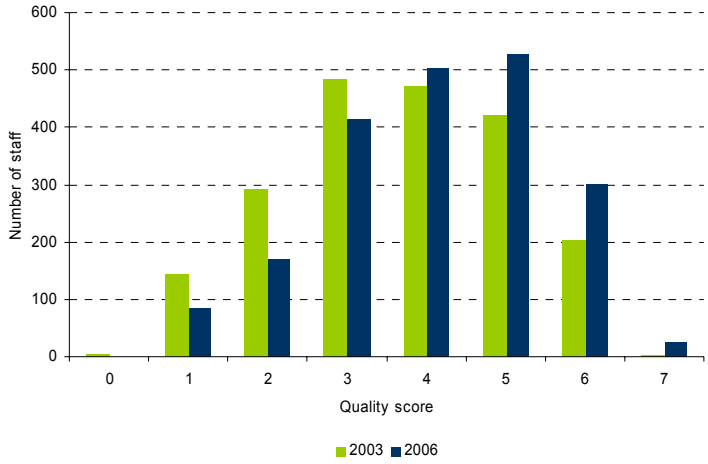
The change in the research component scores allocated to staff can help identify what was driving the improvement in measured research quality. Figure 3 shows that all three average research component scores – research output (RO), peer esteem (PE) and contribution to the research environment (CRE) – showed improvement between 2003 and 2006. However, the highest growth was exhibited by the average PE and CRE scores, with growth of 23 and 24 percent, respectively. This was almost twice the rate of growth in the average RO score (12 percent).

Despite this divergence in growth, the average RO score still remained above the other two average research component scores, although the margin was reduced. In 2006, the average PE score was 93 percent of the average RO score, compared with 85 percent in 2003. Similarly, in 2006, the average CRE score was 91 percent of the average RO score compared with 82 percent in 2003.

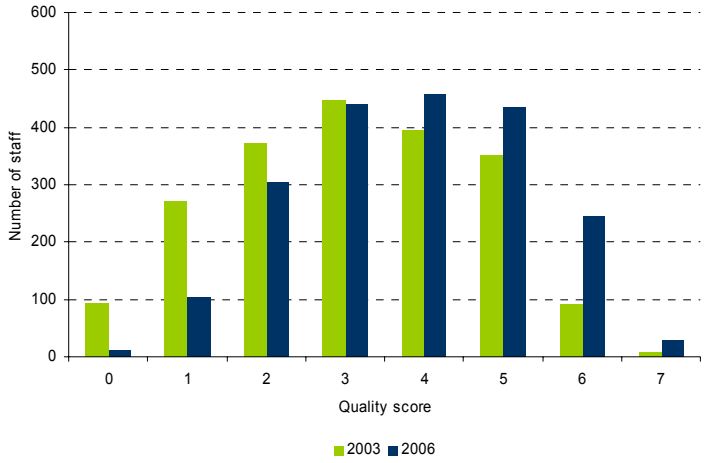
Another way of examining this data is to compare the number of staff allocated to each of the seven possible levels of research component score in 2003 and 2006. Figure 4 shows the distribution of staff by score across each of the three research components. In each research component score, the distribution has shifted to the right indicating improvement in measured research quality. However, the reduction in the number of staff receiving a score of 0 or 1 for their PE score or CRE score in 2006 is particularly noticeable.

Figure 4: Distribution of research component scores

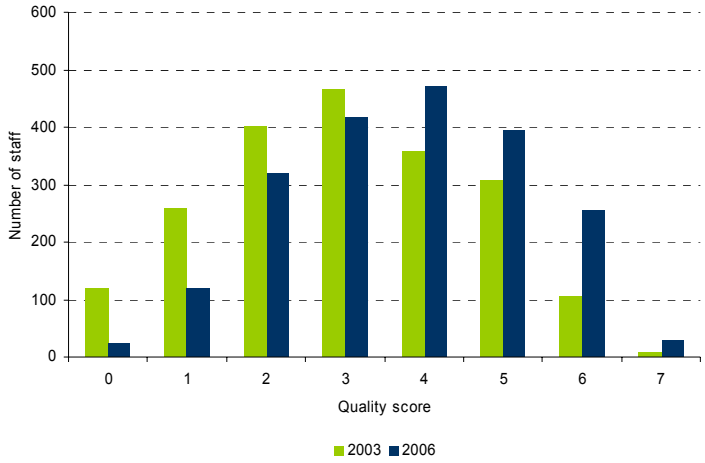
Research output score



Peer esteem score



Contribution to research environment score



5.2 Average change in measured research quality by subgroup

The analysis in this section examines the change in the performance of staff across two dimensions. The first is the percentage of staff in each subgroup whose quality category improved in 2006. The second is the amount of change in the average research component scores of each subgroup between 2003 and 2006. To help illustrate how any changes in measured research performance have affected the relative performance of each subgroup, the actual level of performance in 2003 and 2006 is presented. This includes the distribution of quality categories allocated in 2003 and 2006 and also the average research component scores in 2003 and 2006.

Because the research component scores awarded to staff ranged between a minimum of 0 and a maximum of 7, the maximum amount that any individual staff member's score could improve in 2006 would be 7 points, although a change of this magnitude would be extremely unlikely.

Gender

The percentage of women in the sample whose quality category improved (39 percent) was slightly higher than men (36 percent). This led to a reduction in the disparity between men and women in the quality categories attracting PBRF funding. In 2003, 76 percent of women achieved at least a C quality category, compared with 88 percent for men. By 2006, the gap had closed slightly with 89 percent of women and 96 percent of men achieving at least a C quality category.

Figure 5: Percentage of staff with an improved quality category by gender

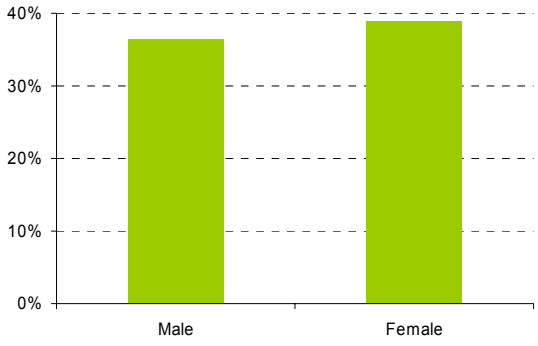
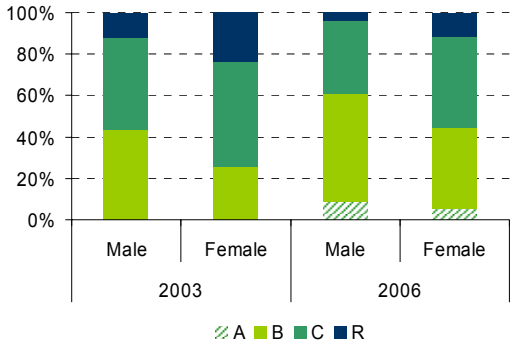


Figure 6: Distribution of quality categories by gender



As can be seen in Figure 7, both men and women achieved a larger increase in their average PE and CRE scores than in RO score. The women in the sample exhibited growth in each of the average research component scores that was around 50 percent higher than men, although this was off a lower base. This has resulted in a reduction in the disparity in average research component scores between men and women observed in the 2003 data.

Figure 7: Change in average research component score by gender

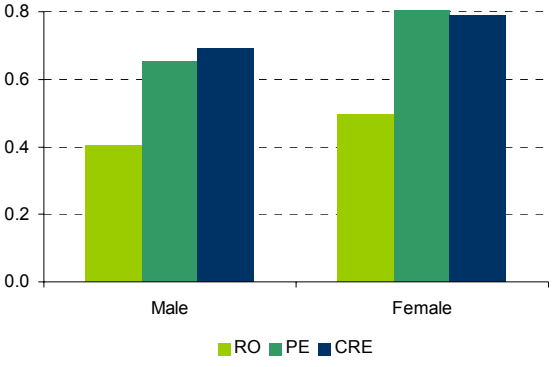
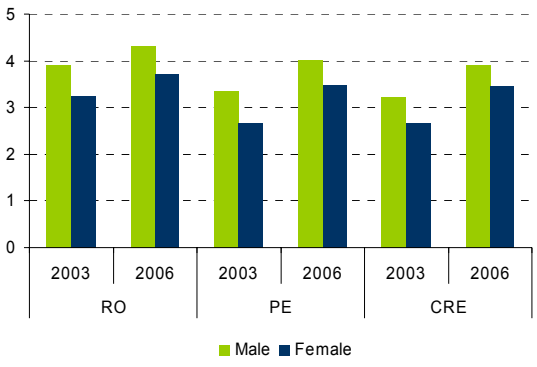


Figure 8: Average research component scores by gender



Ethnic group

Staff in the 'Other' ethnic group had the highest percentage of staff improving their quality category (43 percent), followed by Europeans (37 percent) and Asians (37 percent). The percentage of Maori and Pasifika staff improving their quality category was lower than other ethnic groups, at 26 percent and 31 percent, respectively.

This has resulted in a slight increase in the gap in the proportion of Māori and Pasifika staff achieving at least a B quality category compared with the other ethnic groups (see Figure 10). For example, in 2003, 23 percent of Māori researchers achieved more than a B quality category, compared with 38 percent for Europeans. In 2006, 30 percent of Māori researchers achieved at least a B quality category compared with 56 percent for Europeans.

Figure 9: Percentage of staff with an improved quality category by ethnic group

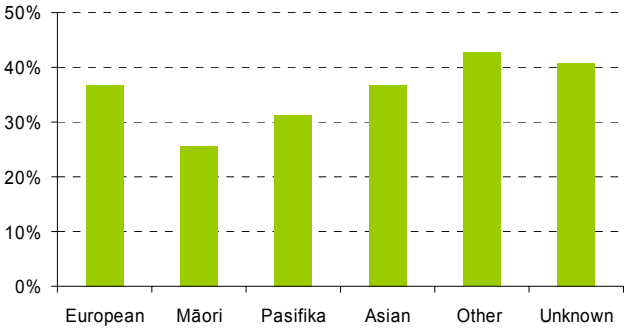
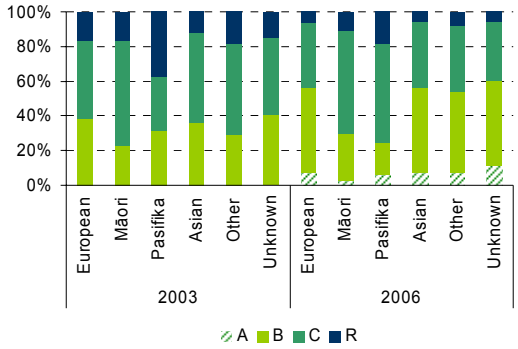


Figure 10: Distribution of quality categories by ethnic group



Generally, the increase in PE and CRE scores was much greater than in RO scores across the various ethnic groups. However, Pasifika staff experienced a slight fall in average CRE score and their change in average RO score was the highest of their three component scores. As shown in Figure 11, the greatest drop in the relative scores of Pasifika staff compared with Europeans was in the CRE score. In 2003, the Pasifika score was 101 percent of the average European score, compared with 89 percent in 2006.

Māori staff exhibited a greater degree of variation in the improvement in their research component scores which had a mixed impact on their scores compared with other ethnic groups. For example, in 2003 the average RO score for Māori was 89 percent that of European staff. In 2006, this fell to 81 percent. However, the average CRE score for Māori was 85 percent of European staff in 2003 but this improved to 88 percent in 2006.

Figure 11: Change in average research component score by ethnic group

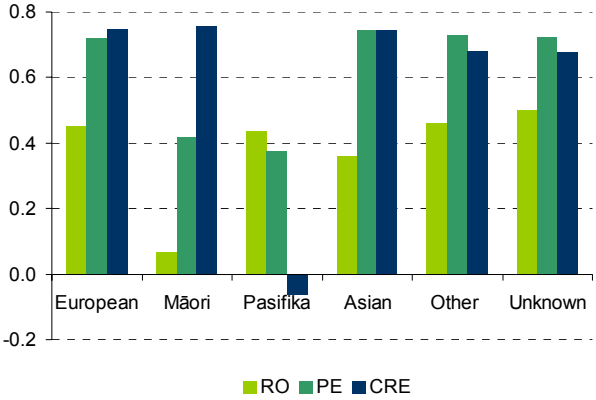
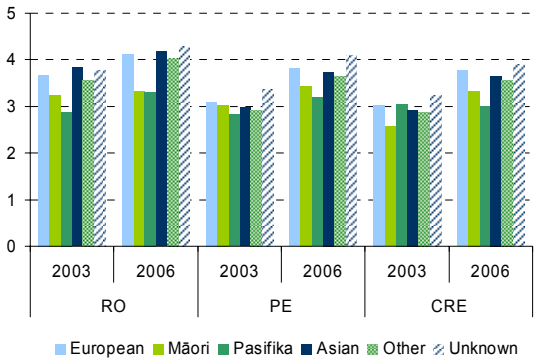


Figure 12: Average research component scores by ethnic group



Position

Staff of higher academic rank had a lower likelihood of improving their quality category. Thirty-one percent of professors improved their quality category, compared with 41 percent of lecturers, 39 percent of senior lecturers and 29 percent of associate professors. This has resulted in a reduction in the disparity in the proportion of staff allocated at least a B quality category between higher and lower academic ranks. In 2003, 86 percent of professors and 11 percent of lecturers in the sample had achieved a B or higher quality category. In 2006, 91 percent of professors and 31 percent of lecturers achieved a B or higher quality category. However, it should be noted that the improvement for lecturers has come off a much lower base than the improvement for professors.

Figure 13: Percentage of staff with an improved quality category by position

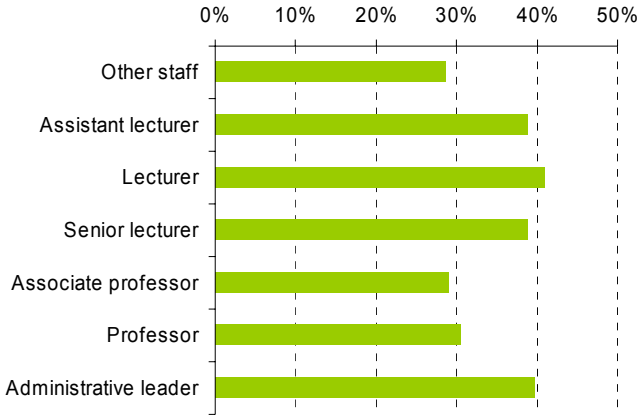
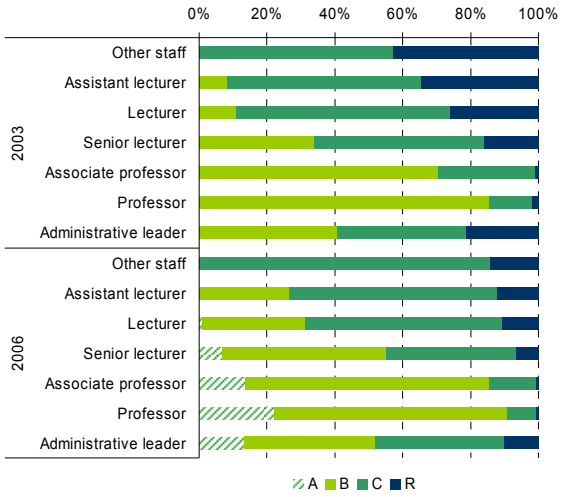


Figure 14: Distribution of quality categories by position



The pattern of higher growth in PE and CRE scores is exhibited by each of the academic positions in Figure 15. Staff in the lower academic ranks achieved the greatest rate of increase in their research component scores, although this was off a lower base. For example, the average RO score for lecturers increased by 16 percent in 2006, compared with 5.3 percent for professors.

Despite the difference in growth rates of the research component scores by position, the average research component scores in 2006 still exhibit a pattern of a higher average research component scores by the higher academic ranks, although the disparities in performance have decreased.

Figure 15: Change in average research component score by position

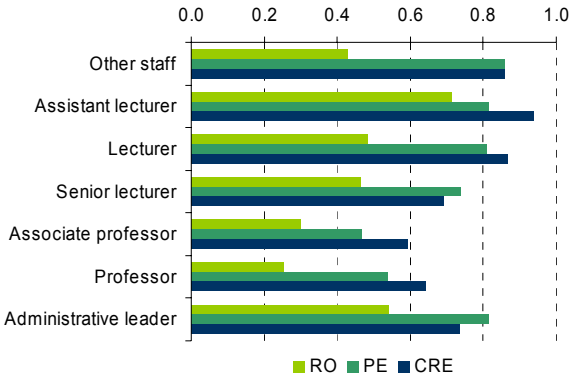
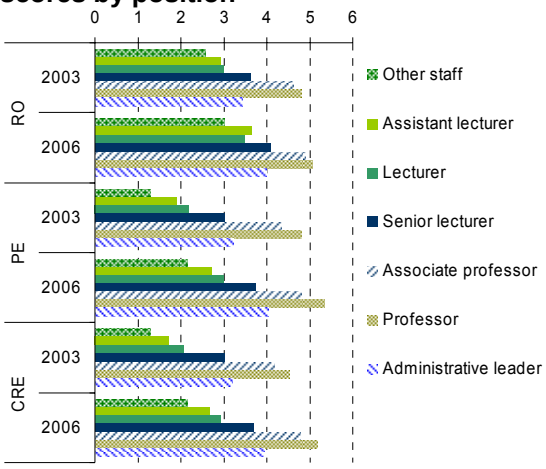


Figure 16: Average research component scores by position



Institution type

Around 41 percent of staff employed at a non-university tertiary education organisation improved their quality category, compared with 37 percent of staff employed at universities. The largest percentage of staff improving their quality categories were staff who changed from a non-university institution in 2003 to a university in 2006 (42 percent).

The gap in the proportion of staff achieving a C or higher quality category between these groups reduced between 2003 and 2006. In 2003, 43 percent of staff employed at a non-university in both Quality Evaluations achieved a C or higher quality category. This compares with 87 percent of staff employed at a university in both Quality Evaluations. In 2006, the proportions were 64 percent and 95 percent, respectively.

Figure 17: Percentage of staff with an improved quality category by institution type

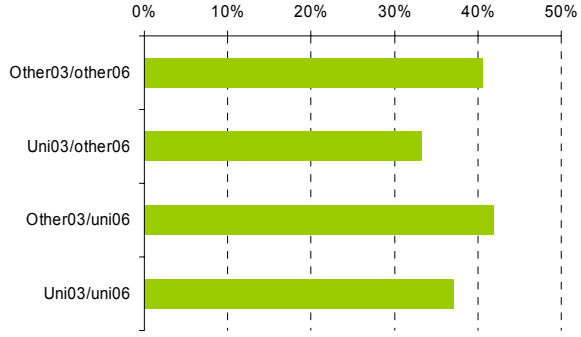
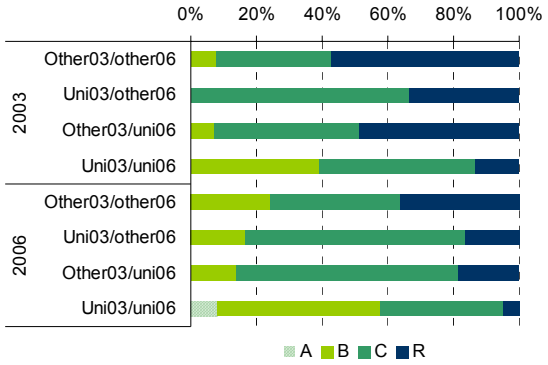


Figure 18: Distribution of quality categories by institution type



In Figure 19, the pattern of a larger increase in the average PE and CRE scores is exhibited by staff employed at either a non-university or a university in both Quality Evaluations. Staff at non-universities in both Quality Evaluations improved their research component scores at a faster rate than staff employed at universities in both Quality Evaluations, although this growth was from a much lower base. This has resulted in a reduction in the gap between the average research component scores for these two groups of researchers, especially in terms of the PE and CRE score. In 2003, the average CRE score of staff at non-university institutions in both Quality Evaluations was 44 percent that of researchers employed at a university in both Quality Evaluations. In 2006, this increased to 57 percent.

Figure 19: Change in average research component score by institution type

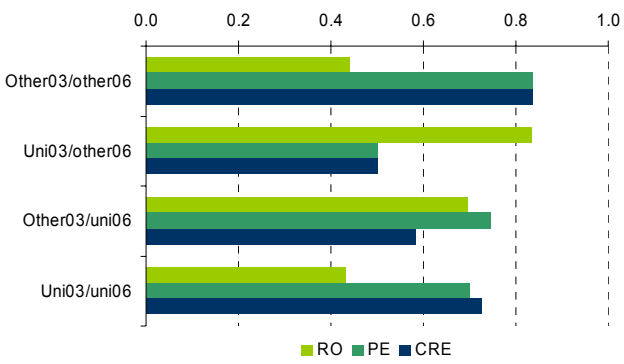
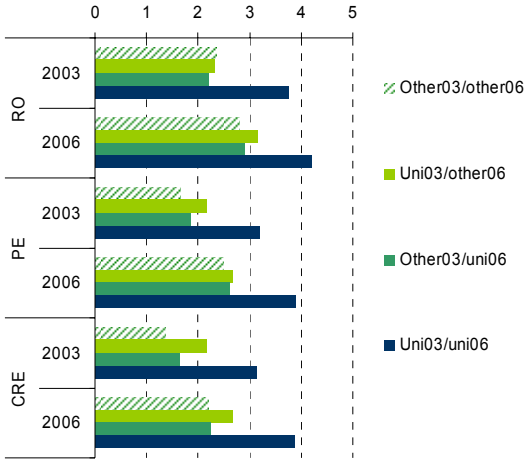


Figure 20: Average research component scores by institution type



Subject panel

Staff in the 'Health' panel had the greatest likelihood of improving their quality category in 2006. Forty-four percent of staff in this area improved their quality category. However, the performance of staff in 'Health' was relatively low in the 2003 Quality Evaluation. The smallest increase was in the 'Māori knowledge and development' panel,[19] where 5.6 percent of staff improved their quality category.[20]

Staff in the 'Physical sciences' panel also had a relatively lower percentage of staff improving their quality category in 2006 (23 percent). However, as can be seen in Figure 22, staff in this area were among the strongest performers in 2003. A similar scenario also applied to staff in the 'Māori knowledge and development' panel, where the small percentage of staff improving their quality category was off a relatively high level of performance in the 2003 Quality Evaluation.

Figure 21: Percentage of staff with an improved quality category by subject panel

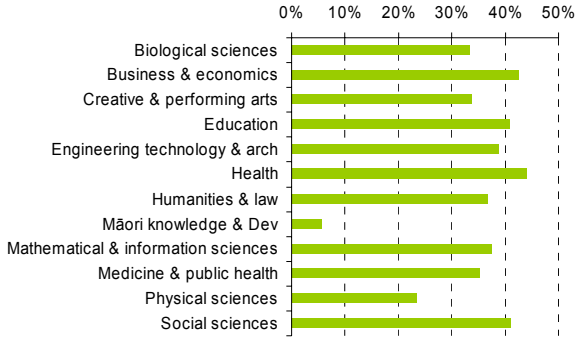
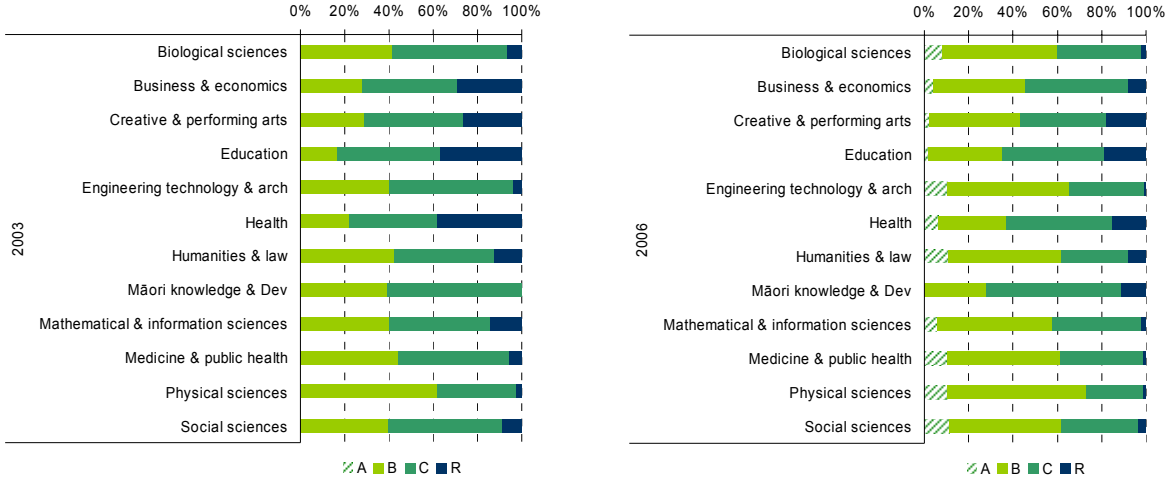


Figure 22: Distribution of quality categories by subject panel 2003 and 2006



[19] A reminder that given the small number of staff in the 'Māori knowledge and development' panel in the study sample (16), this result should be treated with caution.

[20] At the narrow subject level, staff in the 'Veterinary sciences' subject had the highest likelihood of increasing their quality category (65 percent) while those in 'Māori knowledge and development' had the lowest (5.6 percent). The increase in likelihood of attaining a higher quality category for staff in the 'Veterinary science' subject was driven by significant increases in their average PE score (1.7 points) and average CRE score (2.4 points).

As can be seen in Figure 23, the improvement in 'Health' was driven by significant improvements in average PE and CRE scores off a relatively low base. Staff in the 'Māori knowledge and development' panel exhibited a decline in each of their average research component scores, with the greatest decline being in the average RO score.

Staff in the 'Physical sciences' panel also exhibited low average rises in the research component scores. However, as can be seen in Figure 24 to Figure 26 staff in this area had relatively high average scores in 2003 and so therefore had less room for improvement in 2006.

Figure 23: Change in average research component score by subject panel

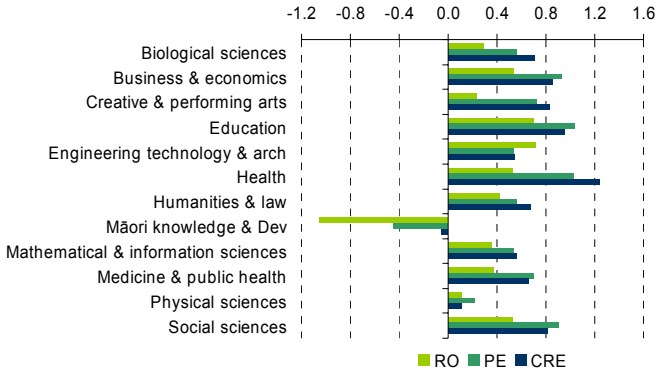


Figure 24: Average RO score by subject panel

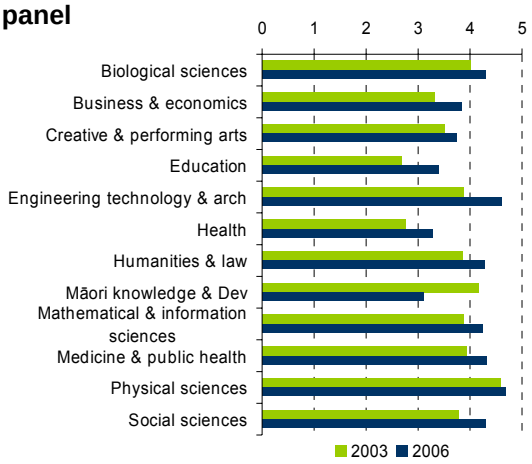


Figure 25: Average PE score by subject panel

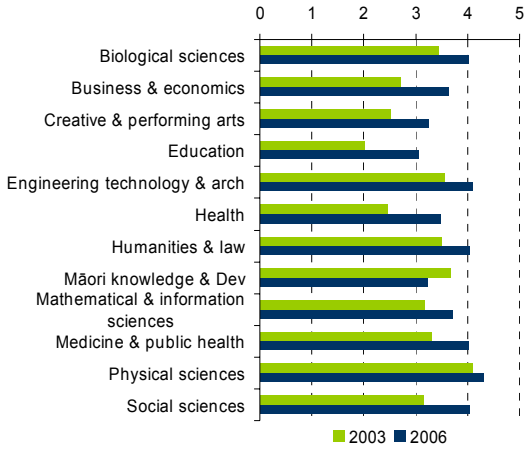
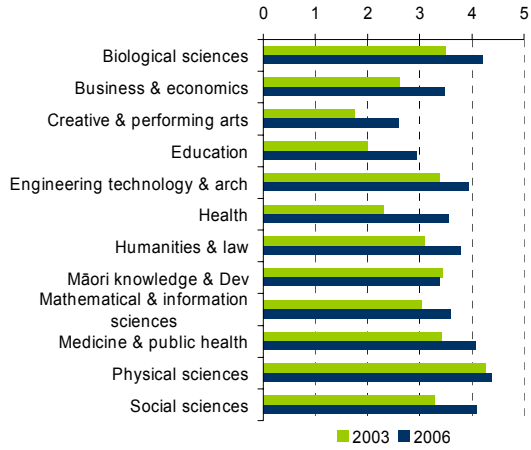


Figure 26: Average CRE score by subject panel

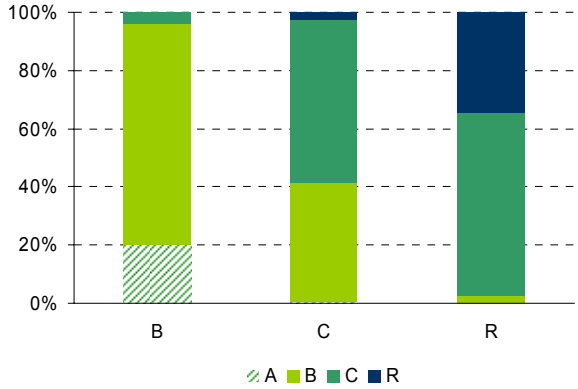


Performance in 2003

A theme throughout the subgroup analysis is that staff who had lower measured research performance in 2003 generally achieved the greatest improvement in quality category in 2006. The statistics illustrated in Figure 27 bear this out, with staff who received lower quality categories in the 2003 Quality Evaluation having a higher likelihood of improving their quality category in 2006. Sixty-six percent of staff who received an R quality category in 2003 were allocated an improved quality category in 2006, compared with 41 percent for staff who received a C and 20 percent for staff who received a B.

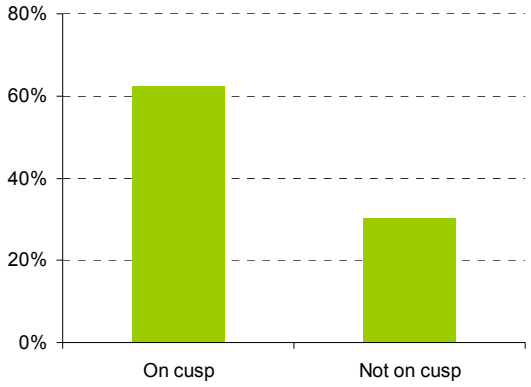
Figure 27: Percentage of staff improving their quality category by 2003 quality category

Figure 28: Distribution of quality categories in 2006 by 2003 quality category



Not surprisingly, those staff who were close to the cusp of a higher quality category in 2003 had a higher likelihood of being allocated a higher quality category in 2006. Staff on the cusp had a 62 percent chance of being allocated a better quality category compared with 30 percent for staff who were not on the cusp.

Figure 29: Percentage of staff with an improved quality category by whether on cusp of higher quality category in 2003



6 Regression analysis results

The previous section used summary statistics to identify the change in the measured research quality by the various characteristics of interest. However, because of possible confounding effects, we cannot be sure how much of the change in performance is associated with a particular characteristic. By using regression analysis we can examine the association between the characteristic of interest and any improvement in measured research quality, while holding all other characteristics constant. This provides a clearer picture of what characteristics are associated with improvement in measured research quality.

6.1 The staff characteristics associated with a higher likelihood of staff receiving an improved quality category

This section presents the results of the logistic regression analysis which identifies the demographic and employment-related characteristics of staff associated with a higher likelihood of staff receiving an improved quality category. The use of regression analysis allows us to control for the assessed performance of staff in 2003, which would appear from the analysis in section 5 to be a key factor associated with change in 2006.

The logistic regression results are presented in Table 1 and include the estimates of the odds ratios, P values (indicating statistical significance) and the Delta P statistic. The statistical significance of the explanatory variables in Table 1 is denoted by a symbol after the odds ratio value. If the odds ratio is followed by an asterisk (*) then the variable is significant at the 5 percent level. If the odds ratio is followed by a two asterisks (**) then the variable is significant at the 1 percent level and if the odds ratio has no asterisk after it then the variable is not statistically significant. For explanatory variables with multiple categories, the statistical significance relates to the difference in likelihood between the category of interest and the reference category.

In this section, the graphs of the odds ratios are presented using a log scale to adjust for the fact they are in ratio form. These graphs also include the 95 percent confidence intervals for the odds ratios estimates. These show the range where we are 95 percent confident the true value of the odds ratio lies within. The confidence intervals are included to show the degree of error associated with the odds ratio estimates and also to allow for a comparison of the statistical significance of two categories of an explanatory variable that do not include the reference category. If the confidence intervals for either of these categories do not contain the estimated value of the odds ratio of the other category then they are significantly different from each other.

The pseudo R^2 value of 0.13 for the logistic regression model suggests that the explanatory power of the model is not particularly strong and other factors outside of the ones included in the model will influence the likelihood of staff receiving a higher quality category. Nevertheless, several explanatory variables in the regression model were found to have a statistically significant association with the likelihood of staff moving to a better quality category.

One of the key factors associated with a higher likelihood of increasing quality category was the level of the quality category allocated to a staff member in the 2003 Quality Evaluation. Figure 30 shows that after controlling for other factors, the lower the quality category allocated to staff in the 2003 Quality Evaluation, the more likely they were to achieve a higher quality category. For example, the odds of a staff member with a B quality category in 2003 being allocated a higher quality category were 90 percent lower than those of a staff member who received an R quality category in 2003. In addition, the odds of a staff member receiving a C quality category in 2003 improving their quality category were around 70 percent lower than for staff who received an R.

Table 1: Logistic regression results

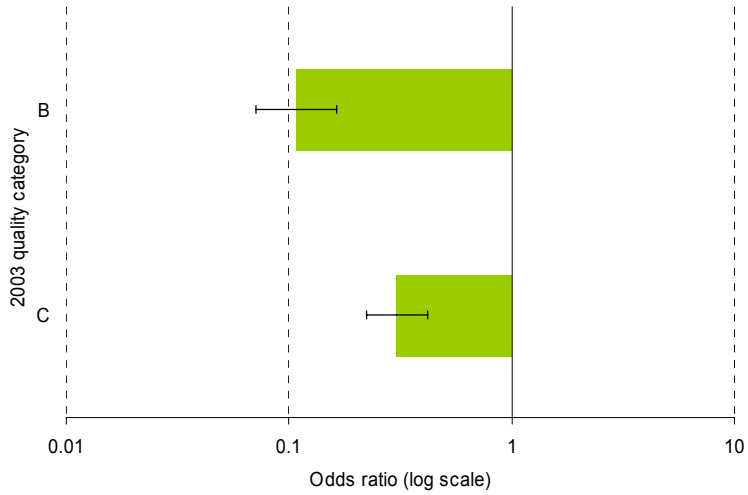
(Dependent variable: improve quality category = 1, otherwise 0)

Characteristic	Category (where applicable)	Odds ratio	P value	Delta P
On the cusp in 2003	Yes	2.41**	0.000	21.6%**
	No		Reference category	
2003 quality category	B	0.11**	0.000	-31.3%**
	C	0.31**	0.000	-21.9%**
	R		Reference category	
Age in 2003		0.97**	0.000	-0.7%**
Gender	Male		Reference category	
	Female	0.86	0.190	-3.4%
Ethnic group	European		Reference category	
	Māori	0.63	0.129	-10.1%
	Pasifika	0.49	0.226	-14.6%
	Asian	0.99	0.953	-0.3%
	Other	1.12	0.530	2.6%
	Unknown	1.18	0.318	3.9%
Change in full-time equivalence		0.99	0.985	-0.2%
Position in 2003	Other staff	0.19*	0.011	-27.1%*
	Assistant lecturer	0.44*	0.014	-16.6%*
	Lecturer	0.59**	0.000	-11.3%**
	Senior lecturer		Reference category	
	Associate professor	1.54*	0.013	10.6%*
	Professor	2.21**	0.001	19.5%**
	Administrative leader	1.64	0.053	12.0%
Change in submitted outputs		1.03**	0.000	0.7%**
Type of institution	Uni03/other06	0.45	0.408	-16.0%
	Other03/uni06	0.80	0.534	-5.2%
	Other03/other06	0.54*	0.024	-13.1%*
	Uni03/uni06		Reference category	
Subject panel in 2003	Biological sciences		Reference category	
	Business & economics	0.99	0.967	-0.2%
	Creative & performing arts	0.99	0.973	-0.2%
	Education	1.18	0.516	3.9%
	Engineering technology & arch	1.36	0.209	7.3%
	Health	0.98	0.943	-0.4%
	Humanities & law	1.17	0.463	3.7%
	Māori knowledge & dev	0.18	0.127	-27.5%
	Mathematical & information sciences	0.98	0.923	-0.5%
	Medicine & public health	1.18	0.462	4.0%
	Physical sciences	0.72	0.245	-7.3%
	Social sciences	1.58*	0.035	11.2%*
Pseudo R^2		0.13		
Log likelihood		-1,162		
N		2,025		

Notes:

1. *, ** represents statistical significance at the 5 percent and 1 percent level, respectively.
2. Delta P statistics represent the percentage point change in predicted probability due to the change in the explanatory variable.

Figure 30: Likelihood of being allocated a higher quality category by level of quality category awarded in 2003 compared with staff allocated an R quality category in 2003



Note: The error bars in the graph show the 95% confidence intervals for the odds ratio estimates.

There is a number of possible explanations for this result. Firstly, staff who received very low quality categories in 2003 would perhaps have a higher motivation to improve on their 2003 performance. The motivation of staff who received an R would be especially strong, given that these staff did not attract any PBRF funding for their tertiary education organisation. Also, some of these staff would be likely to have been identified as new and emerging researchers if that classification had existed at the time and so would also be likely to be improving performance as they gain more experience of research. Secondly, this result suggests that it is harder to move from a B to an A than it is from a C to a B and an R to a C.

The regression results also showed that the odds of a staff member who was close to the cusp of a higher quality category in 2003 improving their quality category were almost 2.5 times that of a staff member not on the cusp.

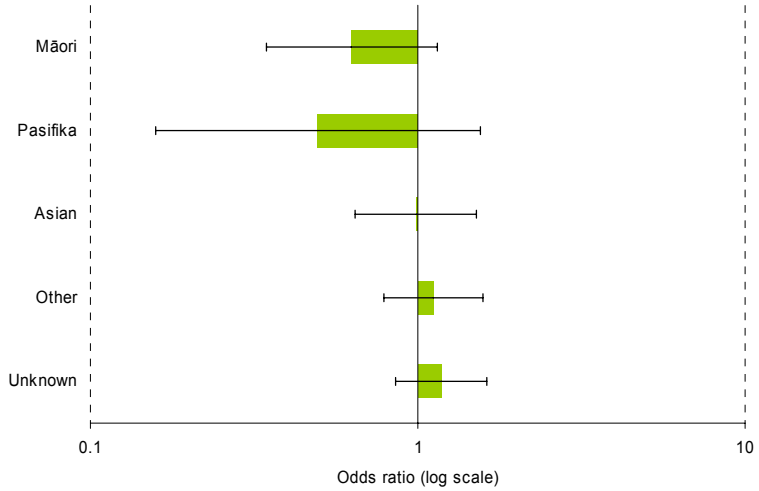
The odds ratio of less than 1 for the age variable in Table 1 indicates that older staff in 2003 had a lower likelihood of being allocated a higher quality category. It is likely that staff nearing the end of their academic careers may be less motivated to improve on their score than younger staff at the beginning of theirs, hence the scale of improvement in scores is lower for older staff. Also, some of the younger staff in the 2003 Quality Evaluation would have been new to the academic profession and so would be expected to show faster rates of improvement than older more experienced staff.

The P value for the gender variable in Table 1 shows that there was no statistically significant difference in the likelihood of achieving a higher quality category for men and women, once other factors (most importantly prior performance) were controlled for. This compares with the results in section 5 that apparently showed a slightly higher likelihood of women receiving a higher quality category. Therefore, it is clear that once other factors were controlled for there was no difference in the likelihood of men and women receiving a higher quality category.

There was also no statistically significant difference in the likelihood of achieving a higher quality category by ethnic group. As can be seen in Figure 31, the 95 percent confidence intervals all include a value of 1, indicating that there is no difference in the likelihood of Europeans and the other various ethnic groups being allocated a higher quality category. This compares with the apparent increasing gap in the performance of the ethnic groups identified in section 5. This shows that once other factors were controlled for there was no difference in the likelihood of staff of any ethnic group improving their quality category.

Similarly, there was also no statistically significant association between a change in the full-time equivalent status of staff and the likelihood of achieving a higher quality category. However, the results in Table 1 show that an increase in submitted research outputs was associated with an increase in the likelihood of being allocated a higher quality category, although the magnitude of the increase was relatively small. In addition, further analysis of the dataset suggested that the positive association between an increase in submitted research outputs and a larger increase in the likelihood of achieving a higher quality category applied mostly to lower academic ranks such as lecturer and senior lecturer.

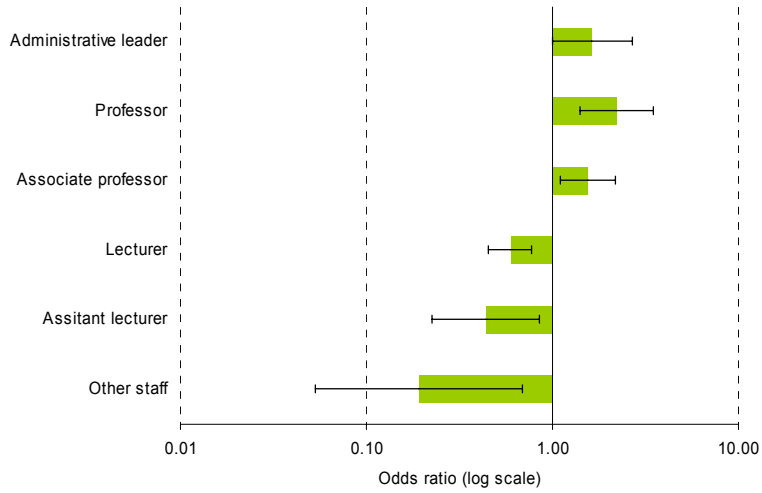
Figure 31: Likelihood of achieving a higher quality category by ethnic group compared with Europeans



Note: The error bars in the graph show the 95% confidence intervals for the odds ratio estimates.

Although the subgroup analysis in section 5 showed that staff of lower academic rank were more likely to improve their quality category, the regression results show that once other factors are held constant, staff in the higher academic ranks (especially professors and associate professors) had a greater likelihood of improving their quality category than staff of lower academic rank. In other words, if a professor and a lecturer received the same quality category in 2003, the professor was more likely to improve their quality category. This may suggest that staff of higher academic rank had a greater degree of motivation to improve their measured research performance and bring it into line with their perceived academic status.

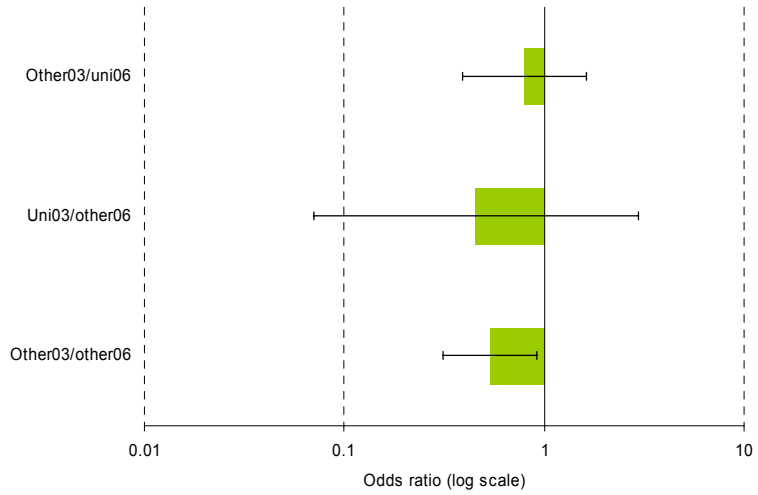
Figure 32: Likelihood of being allocated a higher quality category by position compared with a senior lecturer



Note: The error bars in the graph show the 95% confidence intervals for the odds ratio estimates.

The analysis in section 5 also showed that staff employed at non-university tertiary education organisations in both Quality Evaluations had a greater likelihood of improving their quality category. However, the logistic regression results show that once other factors were controlled for, staff employed at a university in both Quality Evaluations were more likely to improve their quality category. Another way of interpreting this is that staff at a university who received the same quality category as non-university staff would have a higher likelihood of improving their quality category. This suggests that staff at the universities may have had a greater motivation (and perhaps resources) to improve their quality category. This finding is more in line with expectations, given the higher level of research performance required of universities by the Education Act.

Figure 33: Likelihood of being allocated a higher quality category by type of institution compared with staff employed at a university in 2003 and 2006

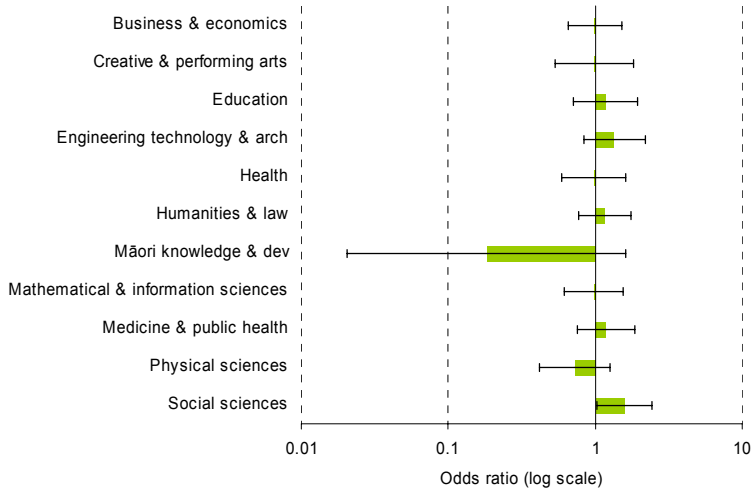


Note: The error bars in the graph show the 95% confidence intervals for the odds ratio estimates.

The subject panel of staff in 2003 was not a particularly significant factor associated with an improved likelihood of being allocated a higher quality category. The 95 percent confidence intervals in Figure 34 show that staff in the 'Social sciences' panel were more likely than the reference category of

'Biological sciences' to improve their quality category, but overall there was little statistical difference in the performance of the subject panels.[21]

Figure 34: Likelihood of being allocated a higher quality category by broad subject panel compared with 'Biological sciences'



Note: The error bars in the graph show the 95% confidence intervals for the odds ratio estimates.

[21] When the model is re-estimated using the narrow subject area, two subjects displayed a higher likelihood of attaining a higher quality category than the majority of the other narrow subject areas. Staff in 'Politics' and 'Veterinary sciences' had a higher likelihood of attaining a higher quality category than staff in the other narrow subject areas.

6.2 The staff characteristics associated with improvement in research component score

This section presents the results of the ordinary least squares regression that identifies demographic and employment-related characteristics of staff associated with change in the research component scores achieved by staff between 2003 and 2006. As was mentioned in section 6.1, the use of regression analysis has the important advantage of controlling for the effect of prior performance on changes in research component scores, allowing for a clearer picture to be obtained of the association between the other explanatory variables used in this analysis and changes in research component scores between 2003 and 2006.

The results of the regression analysis are presented in Table 2 and include the coefficient estimates and P values (which indicate statistical significance). The statistical significance of the explanatory variables in Table 2 is denoted by a symbol after the coefficient value. If the coefficient is followed by an asterisk (*) then the variable is significant at the 5 percent level. If the coefficient is followed by two asterisks (**) then the variable is significant at the 1 percent level and if the coefficient has no symbol after it then the variable is not statistically significant. For explanatory variables with multiple categories, the statistical significance relates to the difference in the change in average research component score between the category of interest and the reference category.

The graphs presented in this section include the 95 percent confidence intervals for the coefficients estimates. These show the range where we are 95 percent confident the true value of the coefficient lies within. The confidence intervals are included to show the degree of uncertainty associated with the coefficient estimates and also to allow for a comparison of the statistical significance of the categories of an explanatory variable that do not include the reference category.

The R^2 values indicate that between 24 and 28 percent of the variation in the change in research component scores was explained by the factors in the model. This suggests that factors outside of the ones included in this model are important determinants of the performance of staff. However, several explanatory variables in the regression model had a statistically significant association with the change in research component scores.

The negative sign of the coefficients for prior performance in Table 2 show that staff who had a lower level of research quality in the 2003 Quality Evaluation achieved larger increases in their research component scores in 2006. For example, for each 1 point decrease in the RO score achieved by staff in the 2003 Quality Evaluation, the RO score allocated to those staff in 2006 improved by 0.35 points. Similarly, a 1 point decrease in the PE score achieved by staff in 2003 was associated with an increase in PE score in 2006 of 0.40 points and a 1 point decrease in CRE score achieved by staff in 2003 was associated with an improvement in CRE score in 2006 of 0.45 points.

The age of staff in 2003 was negatively associated with an increase in measured research quality in 2006. The coefficients for age in Table 2 show that a decrease of 1 year in the age of staff in 2003 was associated with an increase in the RO score received in 2006 of 0.01 points and an increase in PE and CRE scores received in 2006 of 0.02 points.

Table 2: Ordinary least squares regression results

(Dependent variables = change in research output score, change in peer-esteem score and change in contribution to research environment score)

Characteristic	Category (where applicable)	Research component score					
		Research output		Peer esteem		Contribution to research environment	
		Coefficient	P value	Coefficient	P value	Coefficient	P value
2003 research component score		-0.345**	0.000	-0.416**	0.000	-0.448**	0.000
Age In 2003		-0.013**	0.000	-0.016**	0.000	-0.021**	0.000
Gender	Male	Reference category					
	Female	-0.067	0.121	-0.032	0.501	-0.035	0.491
Ethnic group	European	Reference category					
	Māori	-0.247	0.062	-0.079	0.551	0.043	0.775
	Pasifika	-0.377	0.096	-0.487*	0.018	-0.770**	0.002
	Asian	-0.116	0.171	-0.012	0.896	-0.041	0.655
	Other	-0.030	0.672	-0.039	0.654	-0.087	0.316
	Unknown	0.040	0.538	0.043	0.541	-0.064	0.372
Change in full-time equivalence		-0.192	0.222	-0.054	0.725	-0.127	0.487
Position in 2003	Other staff	-0.420	0.150	-0.629**	0.000	-0.610**	0.009
	Assistant lecturer	-0.030	0.823	-0.440**	0.007	-0.446**	0.003
	Lecturer	-0.284**	0.000	-0.372**	0.000	-0.391**	0.000
	Senior lecturer	Reference category					
	Associate professor	0.276**	0.000	0.384**	0.000	0.552**	0.000
	Professor	0.317**	0.000	0.684**	0.000	0.800**	0.000
	Administrative leader	0.195*	0.047	0.355**	0.001	0.400**	0.002
Change in submitted outputs		0.016**	0.000	0.007*	0.035	0.011**	0.001
Type of institution	Uni03/other06	-0.090	0.786	-0.604	0.074	-0.620	0.162
	Other03/uni06	-0.132	0.384	-0.396**	0.006	-0.636**	0.000
	Other03/other06	-0.434**	0.000	-0.477**	0.000	-0.594**	0.000
	Uni03/uni06	Reference category					
Subject panel in 2003	Biological sciences	Reference category					
	Business & economics	0.042	0.615	0.055	0.540	-0.218*	0.022
	Creative & performing arts	0.018	0.894	0.019*	0.885	-0.349*	0.012
	Education	0.187	0.074	0.167	0.111	-0.013	0.906
	Engineering technology & arch	0.377**	0.000	-0.009	0.940	-0.229	0.053
	Health	-0.179	0.072	0.052	0.615	0.017	0.875
	Humanities & law	0.123	0.133	0.023	0.786	-0.177*	0.049
	Māori knowledge & dev	-0.980**	0.000	-0.757**	0.009	-0.708*	0.022
	Mathematical & information sciences	0.012	0.892	-0.161	0.114	-0.353**	0.001
	Medicine & public health	0.061	0.486	0.020	0.835	-0.145	0.163
	Physical sciences	-0.053	0.551	-0.184*	0.047	-0.360**	0.000
	Social sciences	0.221**	0.010	0.269**	0.003	0.090	0.356
Constant		2.215**	0.000	2.733**	0.000	3.164**	0.000
R^2		0.24		0.27		0.28	
N		2,025		2,025		2,025	

Note: *, ** represents statistical significance at the 5 percent and 1 percent level, respectively.

After controlling for other factors, there was no statistically significant difference in the size of increase in research component scores achieved by men and women in 2006. This contrasts with the analysis of the change in average research component scores in section 5, which indicated there was a large difference in the change in scores achieved by men and women. This shows that once other factors are controlled for, most notably prior performance in 2003, the larger improvement in average research component scores by women was not due to gender per se, but was more likely related to the lower performance of women in the 2003 quality evaluation.

There was no statistically significant difference in the increase in RO score achieved by the various ethnic groups, but the increase in PE and CRE scores achieved by Pasifika staff was lower than that of the other ethnic groups.

Figure 35: Change in RO score by ethnic group compared with Europeans

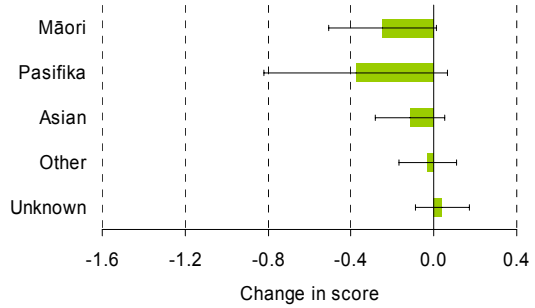


Figure 36: Change in PE score by ethnic group compared with Europeans

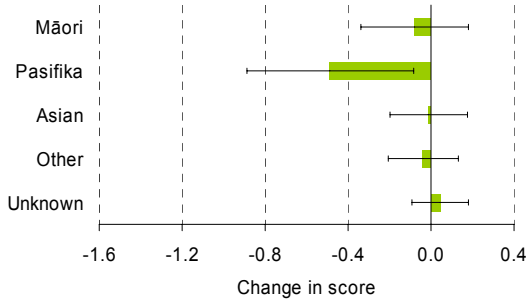
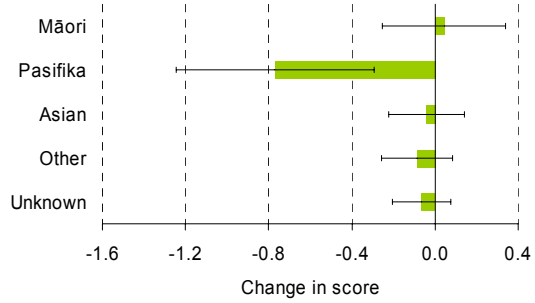


Figure 37: Change in CRE score by ethnic group compared with Europeans



Note: The error bars in the graphs show the 95% confidence intervals for the coefficient estimates.

The change in the full-time equivalent status of staff did not have a statistically significant effect on the increase in research component scores by staff. However, an increase in the number of submitted research outputs was associated with a larger increase in research component scores. An increase of 1 submitted research output boosted the RO score received in 2006 by 0.02 points and the PE and CRE scores received in 2006 by 0.01 points.

However, the scale of the effect of an increase in submitted research outputs is not particularly large. Given that the average increase in submitted research outputs was around 4, an increase of this magnitude in submitted research outputs would only have resulted in an RO score that was on average 0.08 points higher in 2006 than the RO score for a staff member who did not increase the number of their submitted research outputs. In addition, further analysis suggested that it is only in the lecturer and senior lecturer academic ranks that an increased number of research outputs has a positive association with research performance scores.

A higher academic rank was associated with a larger increase in research component scores (see Figures 38 to 40). Professors, associate professors and academic leaders achieved a greater increase in their RO scores than the other academic ranks. The rank of professor was associated with a significantly larger increase in PE and CRE scores than the other academic ranks.

This is the opposite finding to the analysis of the change in average research component scores by position in section 5, which showed that staff of lower academic rank achieved the largest increases in average research component scores. This once again shows how controlling for other factors can help to clarify the nature of the association between a factor of interest (in this case position) and the change in research component score.

Figure 38: Change in RO score by position compared with senior lecturers

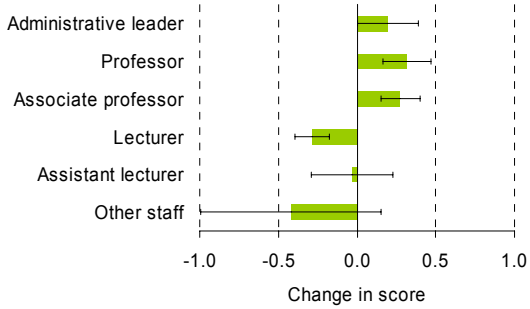


Figure 39: Change in PE score by position compared with senior lecturers

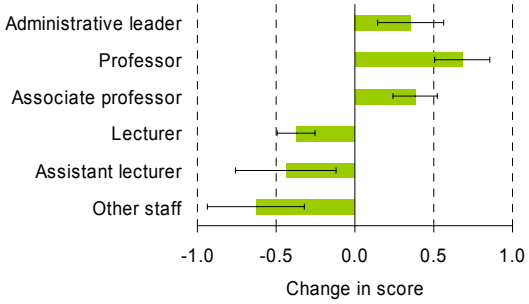
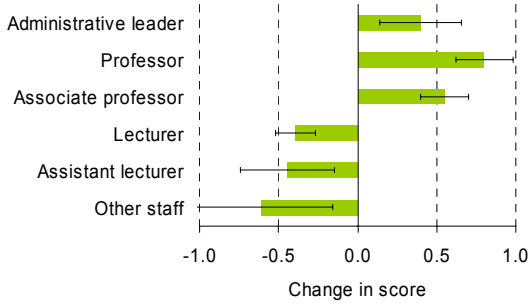


Figure 40: Change in CRE score by position compared with senior lecturers



Note: The error bars in the graphs show the 95% confidence intervals for the coefficient estimates

After controlling for other factors, staff employed at non-universities in both Quality Evaluations achieved a lower increase in research component scores than university staff (see Figures 41 to 43). This once again is the opposite of what was indicated by the raw change in average research component scores in section 5 and shows how controlling for other factors (such as prior performance) can provide a more accurate picture of the nature of the association between the factor of interest and the change in research component scores.

The largest disadvantage was in CRE score, where the increase for staff employed at non-universities in 2003 and 2006 was 0.59 points lower on average than staff employed at a university in 2003 and 2006. Staff who switched from a non-university to a university also exhibited a lower increase in their PE and CRE scores than staff employed at universities in 2003 and 2006.

Figure 41: Change in RO score by insitution type compared with university in both years

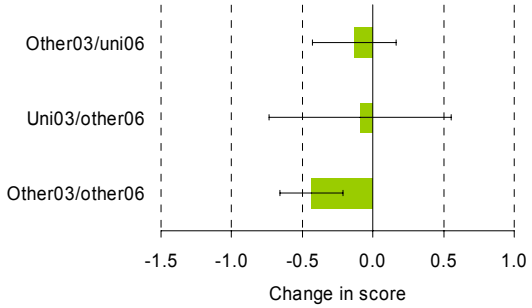


Figure 42: Change in PE score by insitution type compared with university in both years

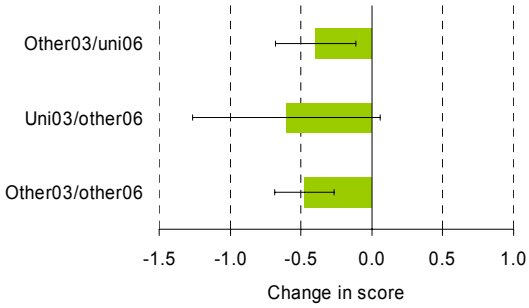
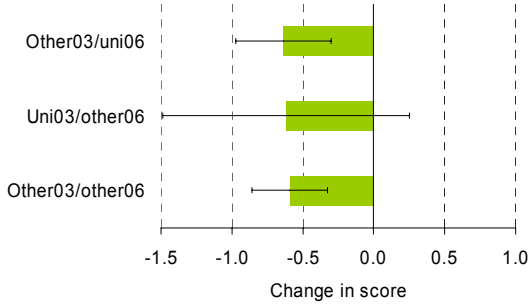


Figure 43: Change in CRE score by insitution type compared with university in both years



Note: The error bars in the graphs show the 95% confidence intervals for the coefficient estimates

Staff in the ‘Engineering’ and ‘Social sciences’ panels achieved a larger increase in RO score than most other panels, while the 'Māori knowledge and development’ panel exhibited a lower increase in RO score than most other subject panels. Staff in the ‘Social sciences’ panel also achieved a greater increase in PE score than most panels.[22]

[22] The analysis of narrow subject areas showed that ‘Politics’ was a strong performer in RO score compared to the other narrow subject areas and ‘Veterinary science’ a strong performer in the CRE score compared to the other narrow subject areas.

Figure 44: Change in RO score by subject panel compared with 'Biological sciences'

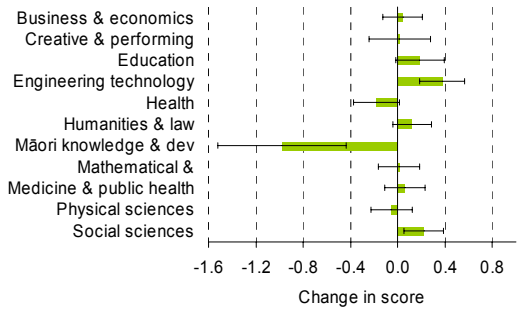


Figure 45: Change in PE score by subject panel compared with 'Biological sciences'

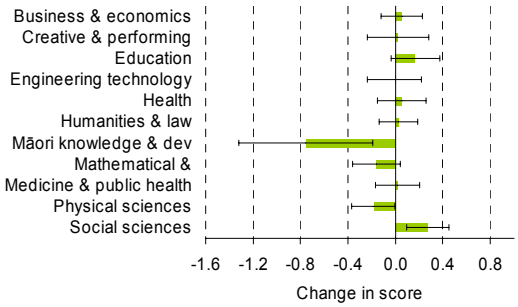
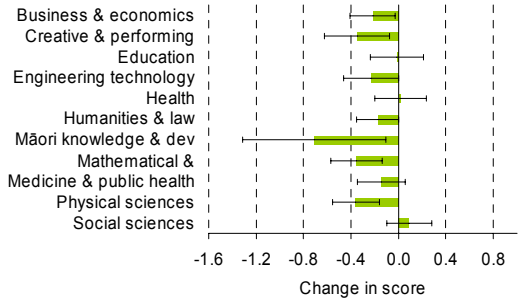


Figure 46: Change in CRE score by subject panel compared with 'Biological sciences'



Note: The error bars in the graphs show the 95% confidence intervals for the coefficient estimates

7 Conclusion

The improvement in measured research performance allocated to the sample of staff used in this study was driven mainly by significant improvement in their peer esteem (PE) and contribution to the research environment (CRE) scores. This greater improvement in the average PE and average CRE score, compared with the average research output (RO) score, suggests that the improvement in measured performance was at least partly due to improved presentation in evidence portfolios given that there is a greater subjective element in the assessment of these dimensions. Nonetheless, there was also a rise in RO score, if to a lesser extent than the other research component scores. Given this research component is potentially less subject to change as a result of improved presentation of evidence portfolios, this provides some evidence of an increase in quality of the research carried out by the staff selected for this study.

The staff in this study who had lower research quality scores in the 2003 Quality Evaluation would appear to have been more successful at improving their measured research quality and to have found it easier to progress up quality categories, compared with staff who achieved higher performance in 2003. In fact, progressing up the quality category scale became progressively less likely as the level of quality category allocated in 2003 increased, which suggests that there is a nonlinear relationship between measured research quality and the quality categories.

However, it is important to remember that this study analysed the performance of a select group of staff, those who participated in the 2003 and 2006 Quality Evaluations, had an evidence portfolio assessed both times and were not identified as new and emerging staff in 2006. Therefore, it should not be assumed that the findings of the analysis apply to all participants in the PBRF Quality Evaluations.

Finally, the way that the nature of association between the increases in measured research quality within the various subgroups examined in this study changed following the application of regression analysis shows that care should be taken when using raw summary statistics to analyse any changes in actual performance. In particular, controlling for the level of measured research performance in 2003 was crucial to separating out the confounding effects of individual staff characteristics on measured research quality.

Appendix: Summary statistics of sample

Table 3: Summary statistics of measured research quality in the 2003 and 2006 PBRF Quality Evaluations

Staff characteristic	Category		Research output score				Peer esteem score				Contribution to research environment score				% of staff with an improved quality category
		N	2003	2006	change	% change	2003	2006	change	% change	2003	2006	change	% change	
All		2,025	3.7	4.1	0.4	12.0%	3.1	3.8	0.7	23%	3.0	3.7	0.7	24.1%	37.3%
Gender	Male	1,281	3.9	4.3	0.4	10.3%	3.4	4.0	0.7	19.5%	3.2	3.9	0.7	21.5%	36.5%
	Female	744	3.2	3.7	0.5	15.4%	2.7	3.5	0.8	30.1%	2.7	3.5	0.8	29.5%	38.8%
Ethnic group	European	1,397	3.7	4.1	0.5	12.3%	3.1	3.8	0.7	23.2%	3.0	3.8	0.7	24.6%	36.8%
	Māori	74	3.3	3.3	0.1	2.1%	3.0	3.4	0.4	13.8%	2.6	3.3	0.8	29.5%	25.7%
	Pasifika	16	2.9	3.3	0.4	15.2%	2.8	3.2	0.4	13.3%	3.1	3.0	-0.1	-2.0%	31.3%
	Asian	125	3.8	4.2	0.4	9.4%	3.0	3.7	0.7	24.9%	2.9	3.7	0.7	25.5%	36.8%
	Other	180	3.6	4.0	0.5	12.9%	2.9	3.6	0.7	25.0%	2.9	3.6	0.7	23.7%	42.8%
	Unknown	233	3.8	4.3	0.5	13.3%	3.4	4.1	0.7	21.5%	3.2	3.9	0.7	21.0%	40.8%
Position	Other staff	7	2.6	3.0	0.4	16.7%	1.3	2.1	0.9	67%	1.3	2.1	0.9	66.7%	28.6%
	Assistant lecturer	49	2.9	3.6	0.7	24.5%	1.9	2.7	0.8	43%	1.7	2.7	0.9	54.1%	38.8%
	Lecturer	519	3.0	3.5	0.5	16.1%	2.2	3.0	0.8	37%	2.1	2.9	0.9	42.1%	41.0%
	Senior lecturer	908	3.6	4.1	0.5	12.8%	3.0	3.8	0.7	24%	3.0	3.7	0.7	23.1%	38.8%
	Associate professor	290	4.6	4.9	0.3	6.5%	4.3	4.8	0.5	11%	4.2	4.8	0.6	14.2%	29.0%
	Professor	154	4.8	5.1	0.3	5.3%	4.8	5.3	0.5	11%	4.5	5.2	0.6	14.2%	30.5%
	Administrative leader	98	3.4	4.0	0.5	15.7%	3.2	4.0	0.8	25%	3.2	3.9	0.7	22.9%	39.8%
Type of institution	Other03/other06	91	2.4	2.8	0.4	18.6%	1.7	2.5	0.8	50.3%	1.4	2.2	0.8	61.3%	40.7%
	Uni03/other06	6	2.3	3.2	0.8	35.7%	2.2	2.7	0.5	23.1%	2.2	2.7	0.5	23.1%	33.3%
	Other03/uni06	43	2.2	2.9	0.7	31.6%	1.9	2.6	0.7	40.0%	1.7	2.2	0.6	35.2%	41.9%
	Uni03/uni06	1,885	3.8	4.2	0.4	11.5%	3.2	3.9	0.7	21.9%	3.1	3.9	0.7	23.2%	37.1%
Panel	Biological sciences	209	4.0	4.3	0.3	7.3%	3.4	4.0	0.6	16.4%	3.5	4.2	0.7	20.1%	33.5%
	Business & economics	251	3.3	3.9	0.5	16.1%	2.7	3.6	0.9	34.5%	2.6	3.5	0.9	32.8%	42.6%
	Creative & performing arts	83	3.5	3.7	0.2	6.5%	2.5	3.2	0.7	28.7%	1.8	2.6	0.8	47.3%	33.7%
	Education	181	2.7	3.4	0.7	26.1%	2.0	3.1	1.0	51.1%	2.0	3.0	1.0	47.7%	40.9%
	Engineering technology & arch	142	3.9	4.6	0.7	18.4%	3.5	4.1	0.5	15.1%	3.4	3.9	0.5	16.0%	38.7%
	Health	141	2.8	3.3	0.5	19.2%	2.5	3.5	1.0	41.9%	2.3	3.6	1.2	53.4%	44.0%
	Humanities & law	291	3.9	4.3	0.4	11.0%	3.5	4.0	0.6	16.2%	3.1	3.8	0.7	21.7%	36.8%
	Māori knowledge & Dev	18	4.2	3.1	-1.1	-25.3%	3.7	3.2	-0.4	-12.1%	3.4	3.4	-0.1	-1.6%	5.6%
	Mathematical & information sciences	166	3.9	4.2	0.4	9.3%	3.2	3.7	0.5	16.9%	3.0	3.6	0.6	18.6%	37.3%
	Medicine & public health	210	3.9	4.3	0.4	9.4%	3.3	4.0	0.7	21.1%	3.4	4.1	0.7	19.2%	35.2%
	Physical sciences	116	4.6	4.7	0.1	2.4%	4.1	4.3	0.2	5.3%	4.3	4.4	0.1	2.6%	23.3%
	Social sciences	217	3.8	4.3	0.5	14.0%	3.1	4.1	0.9	28.9%	3.3	4.1	0.8	24.6%	41.0%

Table 3: Summary statistics of measured research quality in the 2003 and 2006 PBRF Quality Evaluations - continued

Staff characteristic	Category		Research output score				Peer esteem score				Contribution to research environment score				% of staff with an improved quality category
		N	2003	2006	change	% change	2003	2006	change	% change	2003	2006	change	% change	
Subject	Accounting & finance	54	3.5	3.9	0.3	8.9%	2.3	3.6	1.3	58.1%	2.2	3.4	1.3	59.0%	35.2%
	Agriculture & other applied biological sciences	46	3.7	4.1	0.4	10.5%	3.3	4.1	0.7	22.1%	3.4	4.2	0.8	24.5%	41.3%
	Anthropology & archaeology	27	4.2	4.6	0.4	8.8%	3.9	4.5	0.6	15.2%	4.0	4.5	0.5	13.1%	18.5%
	Architecture, design, planning, surveying	41	3.7	4.6	0.9	24.7%	3.7	3.6	-0.1	-2.0%	3.1	3.3	0.2	7.1%	36.6%
	Biomedical	69	4.3	4.5	0.3	6.1%	3.3	4.1	0.7	22.1%	3.6	4.3	0.7	19.4%	34.8%
	Chemistry	47	4.6	4.6	0.0	0.5%	4.0	4.3	0.3	8.0%	4.1	4.4	0.3	6.2%	21.3%
	Clinical medicine	91	3.8	4.3	0.4	11.5%	3.5	4.0	0.5	15.9%	3.3	3.9	0.6	16.8%	36.3%
	Communications	20	3.1	3.9	0.8	24.2%	1.8	3.3	1.5	80.6%	2.3	3.4	1.1	48.9%	45.0%
	Computer science, IT, information sciences	102	3.6	4.1	0.5	14.4%	2.9	3.7	0.8	27.9%	2.9	3.6	0.7	24.1%	46.1%
	Dentistry	10	2.9	3.6	0.7	24.1%	2.4	3.5	1.1	45.8%	2.2	3.2	1.0	45.5%	60.0%
	Design	12	2.4	3.2	0.8	31.0%	1.3	3.1	1.8	131.3%	1.3	2.1	0.8	56.3%	41.7%
	Earth sciences	43	4.5	4.6	0.1	2.6%	4.1	4.2	0.1	3.4%	4.4	4.3	0.0	-1.1%	23.3%
	Ecology, evolution & behaviour	64	3.9	4.2	0.2	6.0%	3.6	4.0	0.4	12.3%	3.5	4.1	0.7	19.0%	32.8%
	Economics	42	4.0	4.2	0.3	6.6%	3.5	4.0	0.5	14.2%	3.5	3.8	0.2	6.0%	26.2%
	Education	181	2.7	3.4	0.7	26.1%	2.0	3.1	1.0	51.1%	2.0	3.0	1.0	47.7%	40.9%
	Engineering & technology	101	4.0	4.6	0.6	16.1%	3.5	4.3	0.8	22.4%	3.5	4.2	0.7	19.1%	39.6%
	English language & literature	41	4.2	4.9	0.7	16.8%	3.6	4.4	0.8	20.9%	3.2	4.1	0.9	26.5%	43.9%
	Foreign Languages & Linguistics	49	3.6	3.8	0.2	5.1%	3.0	3.4	0.3	11.5%	2.8	3.1	0.3	11.8%	30.6%
	History	82	4.0	4.5	0.5	12.5%	3.7	4.0	0.3	8.5%	3.4	3.8	0.5	13.7%	36.6%
	Human geography	23	4.0	4.4	0.3	8.6%	3.4	4.3	0.9	25.6%	3.3	4.3	0.9	27.3%	39.1%
	Law	85	3.5	4.1	0.6	16.8%	3.3	4.2	1.0	29.4%	2.9	3.9	1.0	33.5%	41.2%
	Management	104	3.1	3.8	0.7	22.6%	2.8	3.6	0.8	30.0%	2.7	3.6	0.9	35.4%	50.0%
	Maori knowledge & development	18	4.2	3.1	-1.1	-25.3%	3.7	3.2	-0.4	-12.1%	3.4	3.4	-0.1	-1.6%	5.6%
	Marketing & tourism	51	3.0	3.6	0.6	21.6%	2.3	3.4	1.1	46.2%	2.3	3.1	0.8	34.5%	49.0%
	Molecular, cellular & whole organism biology	99	4.2	4.5	0.3	6.8%	3.4	4.0	0.6	16.5%	3.6	4.3	0.7	18.8%	30.3%
	Music, literary arts & other arts	28	3.9	4.0	0.1	3.7%	3.0	3.8	0.8	25.9%	2.2	3.3	1.1	50.0%	42.9%
	Nursing	29	1.9	2.5	0.6	30.4%	2.2	3.0	0.8	36.5%	2.3	3.4	1.1	50.0%	41.4%
	Other health studies	55	3.1	3.4	0.3	9.9%	2.8	3.6	0.8	29.2%	2.8	3.7	0.8	29.7%	36.4%
	Philosophy	18	4.6	4.5	-0.1	-2.4%	4.2	4.5	0.3	8.0%	3.3	4.4	1.1	33.9%	33.3%
	Physics	26	4.7	5.0	0.3	5.7%	4.3	4.4	0.2	3.6%	4.4	4.5	0.1	2.6%	26.9%
	Political science	28	4.1	4.8	0.6	15.7%	3.1	4.4	1.3	39.8%	2.8	4.1	1.3	44.3%	57.1%
	Psychology	54	4.0	4.5	0.5	13.6%	3.2	4.1	0.9	28.7%	3.6	4.4	0.9	25.0%	46.3%
	Public health	50	3.7	4.1	0.4	10.8%	3.0	3.9	0.9	30.7%	3.3	4.1	0.8	23.5%	34.0%
	Pure and applied mathematics	33	4.7	4.6	-0.1	-1.3%	3.9	4.0	0.1	3.1%	3.7	3.8	0.1	2.5%	18.2%

Table 3: Summary statistics of measured research quality in the 2003 and 2006 PBRF Quality Evaluations - continued

Staff characteristic	Category		Research output score				Peer esteem score				Contribution to research environment score				% of staff with an improved quality category
		N	2003	2006	change	% change	2003	2006	change	% change	2003	2006	change	% change	
Subject cont...	Religious studies & theology	16	4.1	3.9	-0.3	-6.1%	3.6	3.8	0.1	3.4%	3.3	3.6	0.3	7.5%	18.8%
	Sociology, social policy, social work	65	3.4	3.9	0.5	15.8%	3.1	3.8	0.7	23.4%	3.3	3.8	0.6	16.8%	38.5%
	Sport & exercise science	27	2.6	3.4	0.9	33.3%	2.2	3.4	1.2	53.3%	2.0	3.4	1.4	70.4%	40.7%
	Statistics	31	4.0	4.3	0.3	7.3%	3.3	3.3	0.1	2.0%	2.7	3.4	0.6	22.4%	29.0%
	Theatre & dance, film, television	11	3.0	3.7	0.7	24.2%	2.4	3.1	0.7	30.8%	2.1	2.9	0.8	39.1%	36.4%
	Veterinary studies & large animal science	20	3.3	3.8	0.6	16.9%	2.3	4.0	1.7	75.6%	1.6	4.0	2.4	154.8%	65.0%
	Visual arts & crafts	32	3.8	3.7	-0.1	-1.7%	2.6	2.8	0.3	11.0%	1.4	2.0	0.6	44.4%	21.9%
On cusp	On cusp	454													62.3%
	Not on cusp	1,571													30.1%
2003 quality category	B	746													20.1%
	C	950													41.1%
	R	329													65.7%

Table 4: Summary statistics of continuous explanatory variables

Variable	Mean	Standard deviation
Age (2003)	46.2	8.9
Change in submitted research outputs	3.9	7.5

References

- Çinlar, N. & Dowse, J. (2008a) *Human resource trends in the tertiary academic workforce*, Wellington: Tertiary Education Commission.
- Çinlar, N. & Dowse, J. (2008b) *Staffing and performance trends in research subject areas*, Wellington: Tertiary Education Commission.
- Çinlar, N. & Dowse, J. (2008c) *Baseline trends in research in the tertiary education sector*, Wellington: Tertiary Education Commission.
- Ramanathan, R. (1998) *Introductory Econometrics with Applications*, *4th edition*, Florida: Harcourt Brace College Publishers.
- Smart, W. (2005) *What determines the research performance of staff in New Zealand's tertiary education sector?* Wellington: Ministry of Education.
- Smart, W. (2008) *Research par excellence: the factors associated with higher research quality in New Zealand's tertiary education organisations*, Wellington: Ministry of Education.
- Tertiary Education Commission (2007) *Performance-Based Research Fund:* e*valuating research excellence – the 2006 assessment,* Wellington: Tertiary Education Commission.
- White, P. & Grice, J. (2008) *Participation and performance by Māori and Pasifika researchers in the Performance-Based Research Fund, 2003-2006*, Wellington: Tertiary Education Commission.

newzealand.govt.nz