MINISTRY OF EDUCATION
Te Tāhuhu o te Mātauranga

Persistence in doctoral research

Analysing the impact of the PBRF on the retention of doctoral students

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: 64-4-463-8526

Acknowledgements:
The author gratefully acknowledges comments provided by Roger Smyth (Ministry of Education), Bhaskaran Nair (Ministry of Education), Julie Keenan (Ministry of Education), Shona Ramsay (Ministry of Education), four reviewers at the Tertiary Education Commission and Professor Gregor Coster (University of Auckland). The author also gratefully acknowledges David Scott (Ministry of Education) for creating the logistic regression dataset and peer reviewing the report and Alison Lipski, who proof-read this 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

© Crown Copyright
All rights reserved.
All enquiries should be made to the publisher.

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

November 2007

ISBN (Web) 978-0-478-13767-5
ISBN (Print) 978-0-478-13766-8

Persistence in doctoral research

Contents

Figures and tables

1 Summary

Research qualifications – such as doctorates and masters degrees that involve a substantial research project – have two important roles. Firstly, they are the primary means of training New Zealand's research workforce which contributes to the Government's goal of economic transformation. Secondly, undertaking research qualifications in any field provides particular and important skills that can be used in many fields – such as government, industry or management. These skills may for instance, contribute to increases in productivity. Significant resources are invested in this endeavour, so it is important that this investment in research training results in successful outcomes, where postgraduate research students acquire and develop their research skills and successfully complete their studies.

The introduction of the research degree completion (RDC) component of the Performance-Based Research Fund (PBRF) has created a significant financial incentive for participating tertiary institutions to maximise the number of students who complete their postgraduate research-based courses in a timely manner. Although the intention of the RDC component is to raise completion rates of postgraduate research students, the long-term nature of doctoral study means it will be several years before any impact of the PBRF becomes apparent. Therefore, in this study, the retention of doctoral students is used as a proxy for completion – where retention is defined as a doctoral student continuing in study in the following year or completing their doctorate.

To analyse the impact of the PBRF on the likelihood of retention, logistic regression was applied to around 21,000 postgraduate doctoral enrolment and completion records between 2001 and 2006. The advantage of using regression analysis in this case is that it can control for other factors that can influence the likelihood of retention, such as study load, and therefore help isolate any PBRF effect.

Although the explanatory power of the regression model was low, the study found that the impact of the PBRF on the likelihood of retention varied depending on the study load and the age of the doctoral students. The likelihood of retention for younger full-time students was found to have increased slightly since the PBRF was introduced. Although the magnitude of this increase was relatively small, it was statistically significant. This may reflect better pastoral care and/or selection of doctoral students following the introduction of the PBRF.

However, the likelihood of retention for older part-time doctoral students has decreased slightly in the years following the introduction of the PBRF. Although this may be due to more rigorous monitoring of the progress of doctoral students, it may also reflect a tightening in the labour market for highly qualified individuals.

The study also found that Pasifika and Asian students had a slightly lower likelihood of retention than European students. The magnitude of this difference was relatively small, but statistically significant. Students who studied in the fields of science, agriculture and health were more likely to be retained than students studying in the field of society and culture. Once again the magnitude of this difference was small, but statistically significant.

The highest likelihood of retention was exhibited by students in their first year of doctoral studies, with the likelihood of retention falling for students in their later years of doctoral study. The study also found that the gender and residency status of a doctoral student had no statistically significant effect on the likelihood of retention.

Further research that examines the impact of the PBRF on the likelihood of completion of postgraduate research study (including masters-level research-based courses) is planned once this data becomes available.

2 Introduction

The tertiary education sector in New Zealand makes an important contribution towards the government's goal of economic transformation through the training of the research workforce. Significant resources are dedicated to this function, especially in the universities, where the vast majority of postgraduate research students are enrolled. Therefore, it is important that this investment in research training results in successful outcomes, where postgraduate research students acquire and develop their research skills and successfully complete their studies.

This report examines the factors that influence the retention of doctoral students in tertiary study – where retention is defined as a doctoral candidate continuing in study in the next year or completing their doctorate. In doing so, this study analyses what initial impact, if any, the introduction of the Performance-Based Research Fund (PBRF) has had on the retention of doctoral candidates in study.

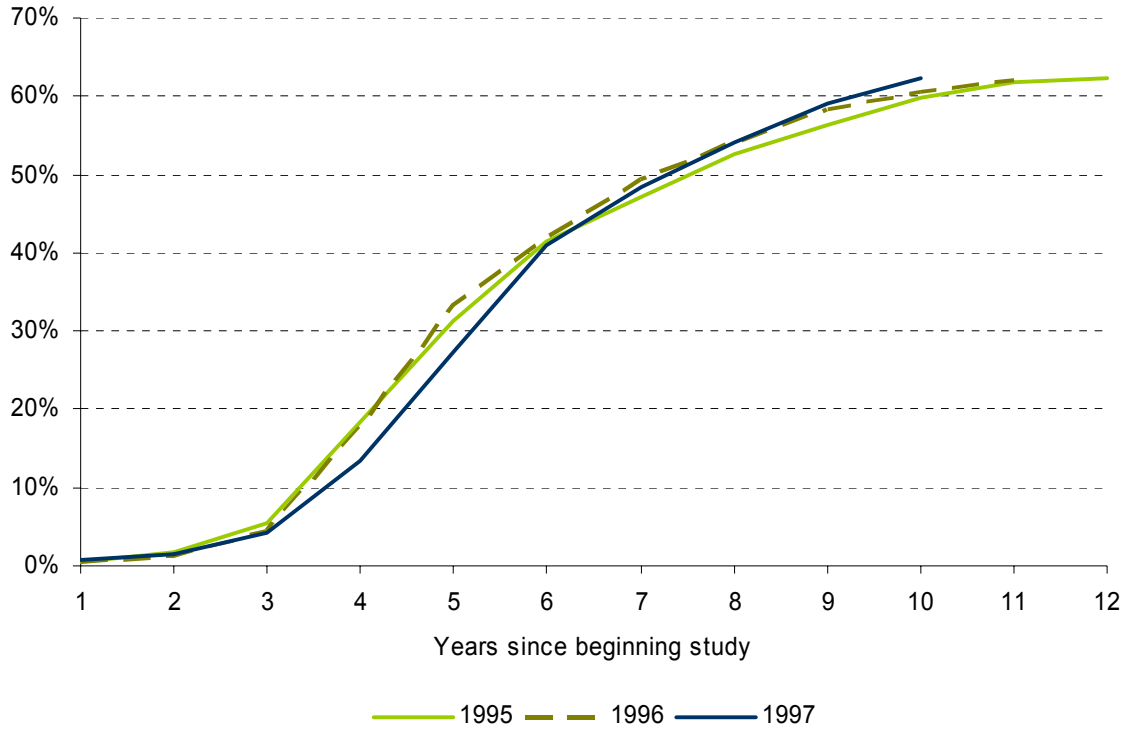
The research degree completion (RDC) component of the PBRF has created a significant financial incentive for tertiary institutions to maximise the number of postgraduate research students who complete their studies and do so in a timely manner.[1] The inclusion of the RDC in the PBRF was partly in response to a perception that the completion rates of postgraduate research students in New Zealand were low (PBRF Working Group, 2002). It was hoped that the completion rate of postgraduate research students would improve as tertiary institutions responded to the changed incentives by reviewing the selection, monitoring and pastoral care of doctoral candidates.

At the time the PBRF Working Group report was released in late 2002, there was no system-wide data available on the completion rates of postgraduate research students. However, the Ministry of Education has since carried out statistical matching of student enrolment and completion records that allows historical completion rates for doctoral students to be estimated. Figure 1 presents the cumulative rate of completion for students starting doctoral studies between 1995 and 1997. The long-term completion rate for doctoral students would appear to be just over 60 percent. This compares with a long-term completion rate of just under 60 percent for bachelors students and around 60 percent for masters students. Therefore, the completion rate of doctoral students is not necessarily as low as may have been perceived by the PBRF Working Group at the time. Also, the completion rates of full-time doctoral students are similar to those in the United Kingdom and Australia (Smart, 2006).

An analysis of the impact of the PBRF on the likelihood of doctoral completion will have to wait until a sufficiently long time-series of completion data has accumulated. Figure 1 shows that the cumulative rate of completion for doctoral students after four years is under 20 percent. If one assumes that the first year that the PBRF may have impacted on the training of doctoral research students was 2003, only around 20 percent of students who started their doctorate in that year would have completed their qualification at the end of 2007, with the completions data not being available until late 2008.

[1] Previously, this funding was allocated solely on the number of domestic enrolments at bachelors-degree and higher level at tertiary institutions.

Figure 1: Cumulative completion rates for doctoral students by starting year



In the interim, analysis of the impact of the PBRF on research training has focused on the retention of postgraduate research students, where retention is used as a proxy for completion. An analysis by Smart (2006) examined trends in first-year retention rates of doctoral students to assess the impact of the RDC component. However, the analysis was limited in that it was descriptive in nature and did not attempt to control for other factors that may influence retention. Also, while the analysis concentrated on the retention rate of students in their first year of doctoral study, it is likely that any changes made by tertiary institutions to the way they train research students would influence all doctoral students, not just students beginning their doctoral studies.

This new report builds upon and extends the previous study by applying logistic regression analysis to the enrolment data of all doctoral students who were studying in the period between 2001 and 2006. The advantage of using regression analysis in this study is that it can control for the impact of other factors, thus allowing for the impact of the PBRF to be examined in isolation. Also, the study examines the retention of all doctoral students from one year to the next, not just doctoral students in their first year, and therefore has a broader scope than the previous study. Although not focusing on the completion rate of doctoral students, an analysis of the factors that impact on the likelihood of retention will at least give an early indication of how the PBRF may be impacting on postgraduate research training.

The RDC component also applies to students undertaking masters-level research courses that are PBRF-eligible. An analysis of the impact of the PBRF on the completion and/or retention of masters students is not part of this study, but is planned for the future once a matched dataset becomes available that identifies those students who are enrolled in PBRF-eligible courses at the masters level.

This report has the following structure. The data used in the analysis and the logistic regression modelling approach are discussed in section 3. In section 4, the results of the logistic regression analysis of the factors associated with the retention of doctoral students are presented. Finally, in section 5, some conclusions are presented.

3 Data, method and the interpretation of predicted probabilities

3.1 Data

The data used in this study is sourced from the administrative records of around 21,000 doctoral enrolments supplied to the Ministry of Education by tertiary institutions. The data includes information on the demographic and study-related characteristics of the students enrolled in doctoral study between 2001 and 2006.

To identify if a student had been retained from one year to the next the enrolment record of a student in one year has been matched to the enrolment and completion record of that student in the following year. For the period prior to 2003 the enrolment and completion records were linked using statistical matching. From 2003 onwards the data has been matched using the National Student Number – a unique life-time identifier introduced in that year.

This change in the method used to link the enrolment and completion records of students does introduce an element of uncertainty in the comparability of the matching between these two periods. However, testing of the matching process suggests that the error in this case is minimal, especially as the matching deals with records in adjacent years.[2]

For this study, the course-level enrolment records of doctoral students have been used to allocate subject areas. Full coverage of the subject area variable is not available until 2001, which limits the analysis in this study to that year onwards.

For the purposes of this study, the logistic regression analysis is restricted to university doctoral enrolments. The universities were the only tertiary institutions offering doctorates in the period prior to the introduction of the PBRF, so to maintain consistency in the dataset the enrolments at other providers have been omitted.[3] Also, due to missing information for the ethnic group of some doctoral students, around 2 percent of doctoral enrolments were excluded from the analysis.

3.2 Method

The report uses binary logistic regression to model the factors that influence the retention of students at the doctoral level. As students are either retained (enrol in the next year or complete their qualification) or not, the dependent variable is binary in nature and takes a value of 1 or 0. Use of ordinary least squares is not appropriate in this case as it will violate the assumption of normality and homoscedasticity of residuals and there is no assurance that the predicted value will lie between 0 and 1 (Ramanathan, 1998). Therefore, logistic regression is used to analyse the impact of the explanatory variables on whether a student returned to study in the following year. In this analysis the enrolment record of each doctoral student in each year is treated as a separate observation. This means that there are around 21,000 observations in the dataset.

Logistic regression applies maximum likelihood estimation after transforming the dependent variable into a logit variable. In this way, logistic regression estimates can be used to calculate the probability of a student being retained in the following year, or not.

The dependent variable in the regression model (*RET*) takes a value of 1 if the doctoral student was also enrolled in the following year or completed their doctoral qualification. If the student did not re-enrol in the following year or complete their qualification, *RET* takes a value of 0.

[2] For more detail on the statistical matching process see Scott (forthcoming).
[3] As a result, around 30 doctoral enrolments at Unitec New Zealand are excluded from the analysis.

While factors such as the financial situation of the student, the strength of the labour market and the grade of the qualification that allowed entry into doctoral studies will impact on the retention of doctoral students, this analysis is limited to the characteristics available in the Ministry of Education dataset used in this report. The explanatory variables in the regression model include demographic and study-related variables and are discussed below.

The age of a student is represented by an explanatory variable with multiple categories (*AGE*). The age groups used in the model were: less than 25, 25 to 29, 30 to 39, 40 to 49 and 50 plus. The reference category used in the model was students aged under 25. This age group was selected as the reference group as younger doctoral students are potentially the most likely to complete their doctorates. They are less likely than older students to have the extra responsibilities of family or work commitments and will have the advantage of a greater continuity of studies from earlier qualifications. A study by Martin et al (2001) of doctoral students in Australia found that the likelihood of completion generally fell with age.

To capture the impact of gender on the likelihood of retention, an explanatory variable (*GENDER*) was included in the regression model. This variable had a value of 1 if the student was a man and 0 if the student was a woman. A previous study by Scott & Smart (2005) on the likelihood of completion at the bachelors level in New Zealand found that women are more likely than men to complete a bachelors degree. However, at the doctoral level, Martin et al (2001) found no statistically significant difference in the likelihood of completion by male and female doctoral students in Australia.

The ethnic group of a student is represented by an explanatory variable with five categories (*ETHNIC*). These categories are: European, Māori, Pasifika, Asian and 'Other'. The reference group used in the regression model is European. This group was chosen as the reference group because they were the largest group in the analysis and the PBRF has a specific funding incentive to encourage completion by Māori and Pasifika students. Therefore, it is useful to have Europeans as the reference group so that the likelihood of retention for these students can be compared with that of Māori and Pasifika students.

An explanatory variable (*INTERN*) is included in the model to represent the residency status of the student. This variable took a value of 0 if the student was a domestic student and 1 if the student was an international student. Because international students generally have a set time period to complete their studies and face considerably higher costs of study than domestic students,[4] it is expected that they may have a higher likelihood of retention.

The study load of a student will probably have a major impact on the likelihood of retention in doctoral studies. Martin et al (2001) found that full-time doctoral students in Australia were more likely to complete their studies than part-time students. The study load of a doctoral student was captured by an explanatory variable (*STUDY_LOAD*) that is derived from the equivalent full-time student (EFTS) load that a student had in that year of enrolment. If a student had an EFTS load of 0.8 or higher they were considered full-time and *STUDY_LOAD* took a value of 0. Otherwise they were considered part-time and *STUDY_LOAD* took a value of 1.

Studies of retention have generally found that students in their first year of tertiary study have the lowest rate of attrition.[5] However, most studies on retention are focused on the undergraduate level, which is likely to be the first engagement a student has with the tertiary education system. Doctoral students, on the other hand, would have already had several years of exposure to the tertiary education system and so would not face the same issues with orientation into the tertiary environment.

[4] Although this changed for new international doctoral students who enrolled from 2006 onwards, who are treated as domestic students and so pay domestic fees.

[5] See for example McGivney (2003), McGinnis (2001) and Williams (1982).

Therefore, it may be that the rate of retention of doctoral students is not at its lowest in the first year but is lower in later years of enrolment. McGivney (2003) lists lack of motivation in long duration programmes of study as a reason for late withdrawal from programmes of study. This is especially pertinent for students who are enrolled in a doctorate, given it is one of the longer programmes of study that students may enrol in. Doctoral students who may begin their studies with high levels of motivation may find their enthusiasm waning after a period of time, thus leading to lower levels of retention at later stages in their programme of study. Also, opportunity costs are higher for doctoral studies and could be perceived by students as magnifying over time.

An explanatory variable with multiple categories (*YEARS_STUDY*) was used to represent the number of years a student had been enrolled in their doctoral studies. The categories used in the model were: one, two, three, four, and five or more years. Therefore, the first year of study is selected as the reference category in the regression model to examine if the likelihood of retention increases or decreases for students in later years of study.

Previous studies on the completion of postgraduate qualifications have found that students in the sciences have a higher likelihood of completion than those in the social sciences/arts areas (for example Martin et al, 2001). A study by the Higher Education Funding Council for England (HEFCE, 2005) suggested that a reason for the sciences having higher rates of completion at the doctoral level was because research in this area was well established and that basic methodological disputes were rare. This compares with some areas in the social sciences and humanities which the study suggested may be less well established and where methodologies may still be disputed.

It is also possible that in sciences a doctorate is the standard terminal qualification. This is less so in other fields, where a masters may be an appropriate end qualification. This may lead to a lower incentive for these students to stay.

An explanatory variable with multiple categories (*SUBJECT*) is used to capture the impact of field of study on the retention of a doctoral student. These categories are based on the two-digit New Zealand Standard Classification of Education (NZSCED) codes and include: science, information technology, engineering, agriculture, health, education, commerce, society and culture, creative arts, and mixed field.[6] The reference category in the regression model is society and culture, which has the highest proportion of doctoral students.

Provider-specific effects are also likely to impact on the likelihood of retention of doctoral students. Martin et al (2001) found that the inclusion of an explanatory variable for institution added significantly to the explanatory power of the analysis of the likelihood of completion by Australian doctoral students. An explanatory variable with multiple categories (*PROVIDER*) is included in the model to capture the effect of provider-specific factors on retention. For confidentiality reasons the values of the coefficients for this variable are not reported in the results.

Choosing a time period when the PBRF might begin to influence the retention of doctoral students is a subjective undertaking. Although the PBRF started allocating funding to tertiary institutions in 2004, providers knew in advance what the measures of performance would entail, suggesting that the first impacts would have been felt before 2004. The concept of linking the funding of tertiary institutions to the number of postgraduate students completing their research courses was initially raised in 2001 by the Tertiary Education Advisory Commission in their recommended funding model for the PBRF (Tertiary Education Advisory Commission, 2001). This was followed by Cabinet's announcement, in May 2002, of the decision to go ahead with the introduction of the PBRF. The details of how the PBRF would work operationally were then outlined in December 2002.[7]

[6] There were no enrolments in the food and hospitality area.
[7] See PBRF Working Group (2002).

In this study it is assumed that the impact of the PBRF would have begun following the confirmation of the operational details of the PBRF at the end of 2002. Therefore, it is assumed that any impact of the PBRF would first impact on students that were enrolled from 2003 onwards. Individual staff members may or may not have responded to the introduction of the PBRF, but the response at the institutional level has been significant as the universities reorganised their research postgraduate operations in anticipation of its implementation.

Coster (2006) noted that the response of the universities to the PBRF has taken the form of developing strategic plans to increase postgraduate student completions and improve the quality of postgraduate supervision. Coster (2006) also identifies that the research culture at the universities has been addressed to place greater emphasis on the importance of postgraduate research and its outcomes. In addition, the study found that there has also been an increase in investment in financial assistance and incentives for postgraduate research students. For example, the University of Auckland began offering doctoral students a financial reward if they completed their doctoral studies in a timely fashion (University of Auckland, 2004).

In the logistic regression model, the impact of the PBRF on the retention of doctoral students is captured using an explanatory variable with multiple categories (*PBRF*) which assumes that the impact of the PBRF can change over time. The categories in the model are *pre-PBRF* (representing students who were enrolled in 2001 and 2002), $PBRF_{03}$ (if a student was enrolled in 2003), $PBRF_{04}$ (if a student was enrolled in 2004) and $PBRF_{05}$ (if a student was enrolled in 2005).

Because the *PBRF* variable is time-based, the impact of other events that occur over time may be captured by this variable. For example, any normal, ongoing reviews that may take place annually at the universities to improve policies and procedures for postgraduate training may also be reflected in the PBRF variable. In other words, if there is an improvement in the retention rates of doctoral students following the introduction of the PBRF, part of this may be due to the normal operational processes of the universities.[8] Nevertheless, the RDC element of the PBRF is arguably the single greatest factor that would have influenced the behaviour of the universities in the 2003 to 2006 time period.

The logistic regression model is presented in equation 1, where the dependent variable is the log of the odds of being retained:

$$(1)\quad \ln[RET_i/(1-RET_i)] = \beta_1 + \beta_2\, AGE_i + \beta_3\, GENDER_i + \beta_4\, ETHNIC_i + \beta_5\, INTERN_i + \beta_6\, SUBJECT_i + \beta_7\, PART\_TIME_i + \beta_8\, YEARS\_STUDY_i + \beta_9\, PBRF_i + \beta_{10}\, PROVIDER_i + \mu$$

Where μ is an error term and $i = 1$ to n observations.

As retention rates for New Zealand doctoral students are relatively high (over 90 percent), non-retention could be considered a 'rare' event. This can sometimes result in standard logistic regression overestimating the probability of the retention of students (King & Zheng, 2001). To test for this, the logistic regression equation was also estimated using an algorithm written by King, Tomz & Zheng (1999) that corrects for this bias, but there was little difference in the coefficients and standard errors obtained from using standard logistic regression. The model was also estimated using Poisson regression, but this also produced similar results to the basic logistic regression model. Therefore, the standard logistic regression results have been used in this report. The results of the regression modelling are presented in logit form in Table 1 in Appendix A. However, to aid interpretation, the results of the logistic regression modelling in this study are expressed in the form of predicted probabilities. These predicted probabilities are derived from the logistic regression equation and

[8] To explore this possibility, an alternate specification of the regression model is presented in Appendix B. The alternate specification includes a time trend variable to control for these other factors that may impact on the likelihood of completion.

represent the probability of a student from a selected reference group successfully being retained from one year to the next.

3.3 Interpreting predicted probabilities

Predicted probabilities are generated by inserting the characteristics of the reference group into the regression equation and then converting the log of the odds to a probability. Then the characteristic of interest can be altered and the predicted probabilities will show how the change in this characteristic will impact on the probability of that student being retained, holding all other factors constant.

As predicted probabilities differ depending on the selected reference group, caution should be used when examining the magnitude of the change as it will vary for different reference groups. However, the direction and statistical significance of the change will remain the same. In this study, the following reference group was selected to generate the predicted probabilities described in this report:[9]

Age group:	<25
Ethnic group:	European
Subject:	Society and culture
Year of doctoral study:	First
Study load:	Full-time
Provider:	An average New Zealand university

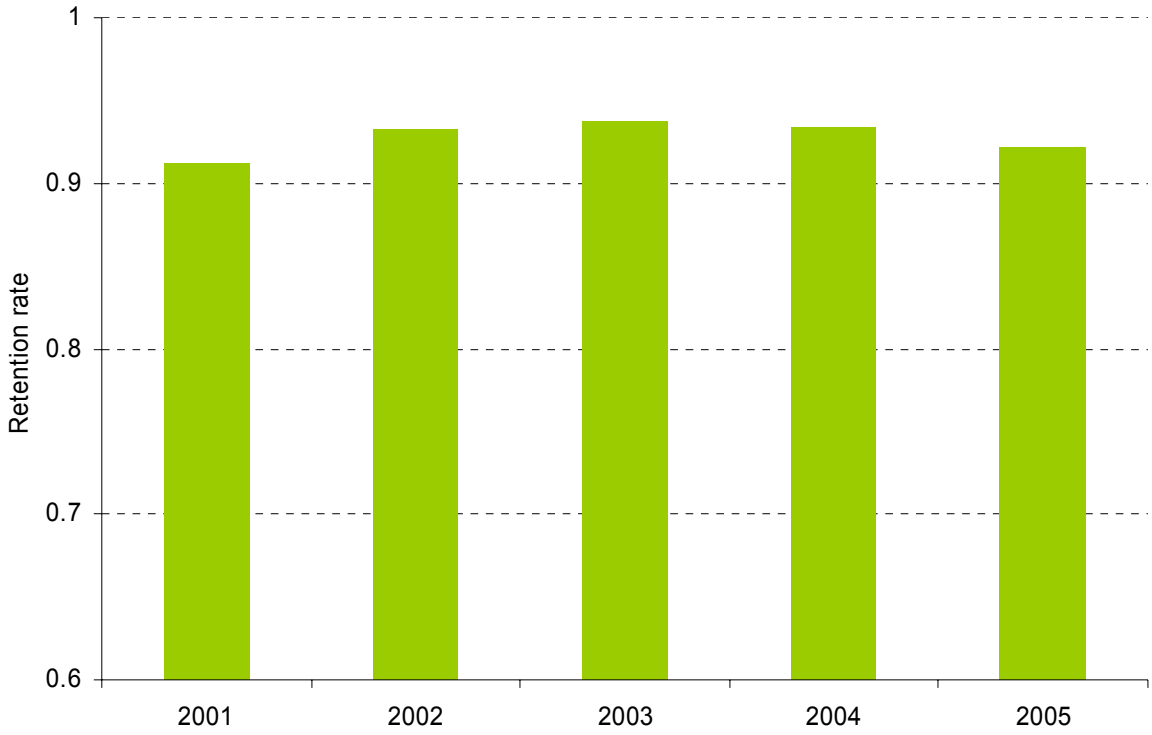
[9] The rationale for selecting these categories was discussed in the preceding text.

4 Retention of doctoral students

4.1 Retention rates for doctoral students

The actual retention rate of doctoral students from one year to the next was over 90 percent between 2001 and 2005. Figure 2 shows that for students enrolled in 2001, 91 percent were retained in the following year. The retention rate then increased to a peak of 94 percent for students enrolled in 2003, before declining slightly to 92 percent in 2005.

Figure 2: Retention rate of doctoral students 2001-2005



This might suggest that the retention rate of doctoral students has been falling since the introduction of the PBRF. However, the impact of confounding factors can mask the impact of policies, such as the PBRF, on the retention rate of doctoral students. The purpose of using regression analysis is that it enables us to isolate the impact of each variable on the likelihood of retention. The results of the regression analysis are explored in the following section.

4.2 Logistic regression results[10]

4.2.1 Summary statistics

Table 1 in Appendix A provides summary information on the characteristics of the doctoral students in the regression dataset. Also, the raw retention rates for the various groups in the model are presented.

A comparison of the retained and non-retained groups used in the logistic regression dataset in Table 1 shows some differences in the make-up of the two groups. Retained doctoral students tend to be younger, full-time and in their first year of study. Also, a higher proportion of science students is retained compared with other fields, while in the non-retained group there is a higher proportion of students in society and culture and in education.

The raw retention rates for the various groups are also presented in Table 1. As can be seen, there is little difference in the retention rate of men and women and of international and domestic students. The retention rate declines for older students, is also higher for full-time students and is highest for students studying in the sciences.

4.2.2 Goodness of fit

Various model specifications were trialled in the regression modelling. It was found that neither the gender[11] nor the residency status[12] of doctoral students had a statistically significant impact on retention and so these variables were dropped from the model. The modelling also showed that there were significant interaction effects between the PBRF and the age of a doctoral student and between the PBRF and the study load of a doctoral student. In other words, the impact of the PBRF on the likelihood of retention varied depending on the age and the study load of the doctoral student.

The goodness of fit of the model was low with a pseudo R^2 value of around 6 percent. This would suggest that factors outside of the ones included in the regression model are having a major impact on the likelihood of retention. This is not a surprising result given the limited amount of information available through the use of administrative datasets and the wide range of factors that can potentially influence decisions to remain in study.[13]

Nevertheless, the regression modelling showed that a number of individual factors had a statistically significant impact on the retention of students.[14] These factors are discussed in the following sections, beginning with the impact of the PBRF on the likelihood of retention of doctoral students.

[10] The logit coefficients of the explanatory variables and their associated standard errors are presented in Table 2 in Appendix A. Model 1 in the table presents the estimates from the main effects model, while Model 2 includes both main effects and interaction effects. It is the results from Model 2 that are reported in this study. The regression results were obtained using STATA v 8.2 (Statacorp, 2005).

[11] A study on the factors influencing the completion of doctoral degrees in the United Kingdom also found that completion rates of men and women were very similar (HEFCE, 2005).

[12] This compares with the finding in the HEFCE (2005) study that international students in the United Kingdom were more likely to complete their doctoral qualification than domestic students. However, modelling of the likelihood of *completion* of doctoral degrees in New Zealand indicates that international students are more likely to complete their doctoral qualification than domestic students – a finding that aligns with the United Kingdom study.

[13] Such as the financial circumstances of the student.

[14] The model χ^2 statistic was also significant at the 1 percent level.

4.2.3 *The impact of the PBRF on retention*

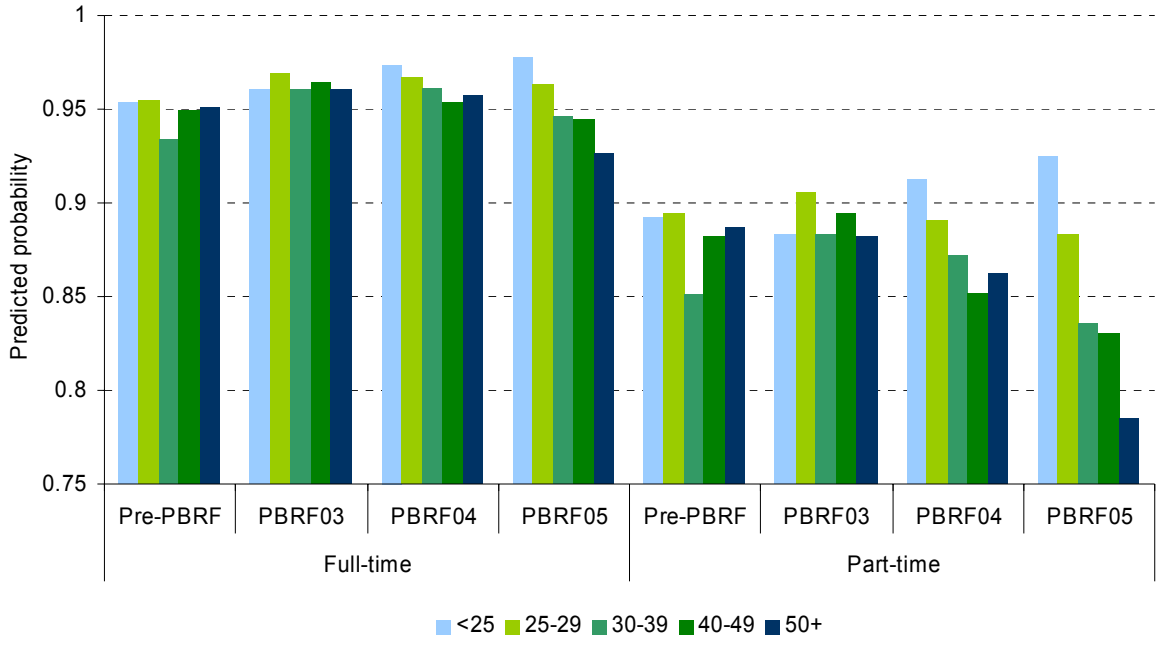
The results of the logistic regression analysis suggest that the introduction of the PBRF has been associated with a small, but statistically significant, impact on the likelihood of retention of doctoral students in New Zealand. Interestingly, the modelling suggests that the impact of the PBRF has varied, depending on the age and study load of the student. It would also appear that the impact of the PBRF is increasing over time.[15]

The effect of the PBRF on the likelihood of retention can best be illustrated by showing the predicted probabilities for doctoral students by age group and study load over time for the selected reference group (see Figure 3). Figure 3 shows that the likelihood of retention for younger doctoral students, especially those aged under 25, has increased in the time period following the introduction of the PBRF for the selected reference group. Prior to the PBRF, the likelihood of retention for full-time students in this age group was 0.95 - by 2005 this had risen to 0.98. For part-time students in this age group, the predicted probabilities were 0.89 and 0.92, respectively.

This compares with a decrease in the likelihood of retention for older students, especially those aged over 40, as time has gone on.[16] Prior to the PBRF, the predicted probability of retention for a full-time student aged 50 and over was 0.95 - by 2005 this had fallen to 0.93. For a part-time student in this age group, the predicted probabilities were 0.89 and 0.78, respectively.

The results of the regression modelling also indicate that although part-time students have always had a lower likelihood of retention than full-time doctoral students, this gap has widened in the period following the introduction of the PBRF.[17]

Figure 3: Predicted probability of retention by age group, study load and PBRF status



Note: The reference group is a European doctoral student, studying society and culture, in their first year of study and at an average New Zealand university.

[15] Although the coefficients of the PBRF variables in the main effects model (Model 1 in Table 1) would suggest that the impact of the PBRF is waning over time, this is due to the fact that the increasing rate of retention for younger doctoral students is being offset by the declining rate of retention for older doctoral students.

[16] This is indicated by the interaction variables $PBRF_{05}$ x *(40-49)* and $PBRF_{05}$ x *(50+)* being statistically significant in Model 2.

[17] This is indicated by the interaction variables $PBRF_{04}$ x *part-time* and $PBRF_{05}$ x *part-time* being statistically significant in Model 2.

The trends exhibited in Figure 3 may reflect improvements made by the universities to the pastoral care and support of doctoral students as a result of the PBRF, which in turn have resulted in improved retention rates of younger full-time doctoral students. The decline in the retention rate of older part-time students, on the other hand, may reflect the effect of tighter monitoring of the progress of doctoral students following the introduction of the PBRF.

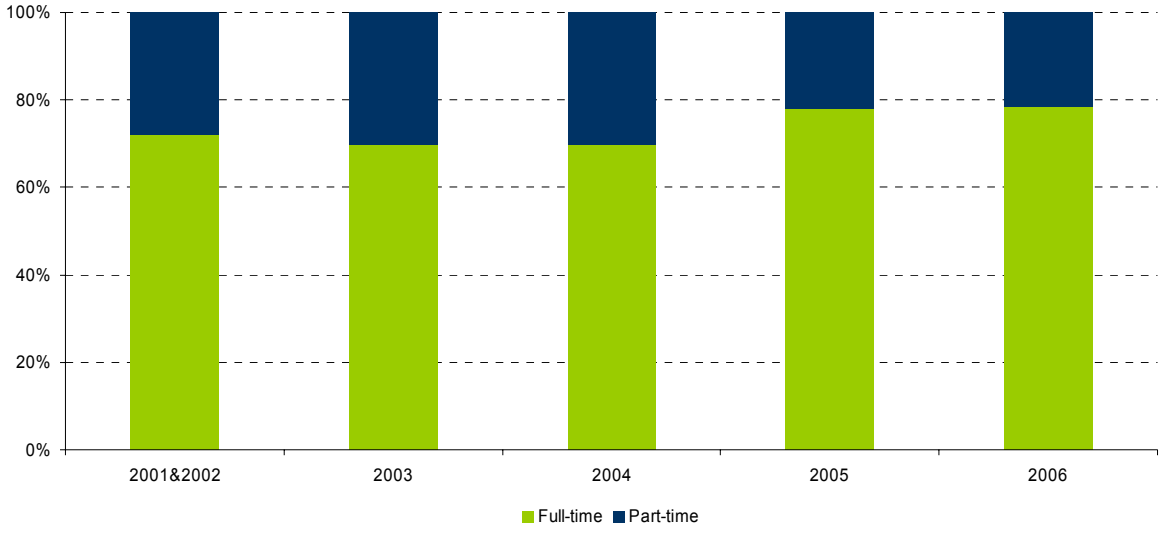
There are other potential causes of the trends exhibited in Figure 3. For example, the lower likelihood of retention for older part-time students may also be a result of the relatively tight nature of the labour market. For example, the unemployment rate for people with a bachelors or higher qualification has fallen from 3.6 percent in 2002 to 2.1 percent in 2006 (Ministry of Education, 2007).

The impact of the labour market on the environment for doctoral students may also be indicated by the proportion of starting doctoral students that are part-time. A decrease in this proportion over time may indicate that people who may have previously considered part-time study may choose instead to pursue employment opportunities, due to the increase in the opportunity cost of pursuing doctoral study.

Because of a change to the funding of international doctoral students in 2006 (where they are now funded at the same rate as domestic doctoral students and are subject to domestic student fees) there has been a significant increase in the number of international students starting a doctorate. Therefore, the analysis of the study load of starting students is restricted to domestic doctoral students to avoid any clouding of trends.

Figure 4 presents the proportion of full-time and part-time starting domestic doctoral students from 2001 and 2002 to 2006. It shows that the proportion of students studying part-time (with fewer than 0.8 EFTS) has decreased from around 30 percent in 2004 to around 22 percent in 2005 and 2006.

Figure 4: Distribution of starting domestic doctoral students by study load



Therefore, the decrease in the likelihood of older part-time doctoral students being retained may be at least a partial reflection of students discontinuing with study in order to focus on increased job opportunities or due to increased conflict between work and study commitments.

The impact of other factors, such as normal ongoing improvements to the operation of postgraduate research training at the universities not related to the PBRF, is explored in an alternative specification of the regression model in Appendix B.

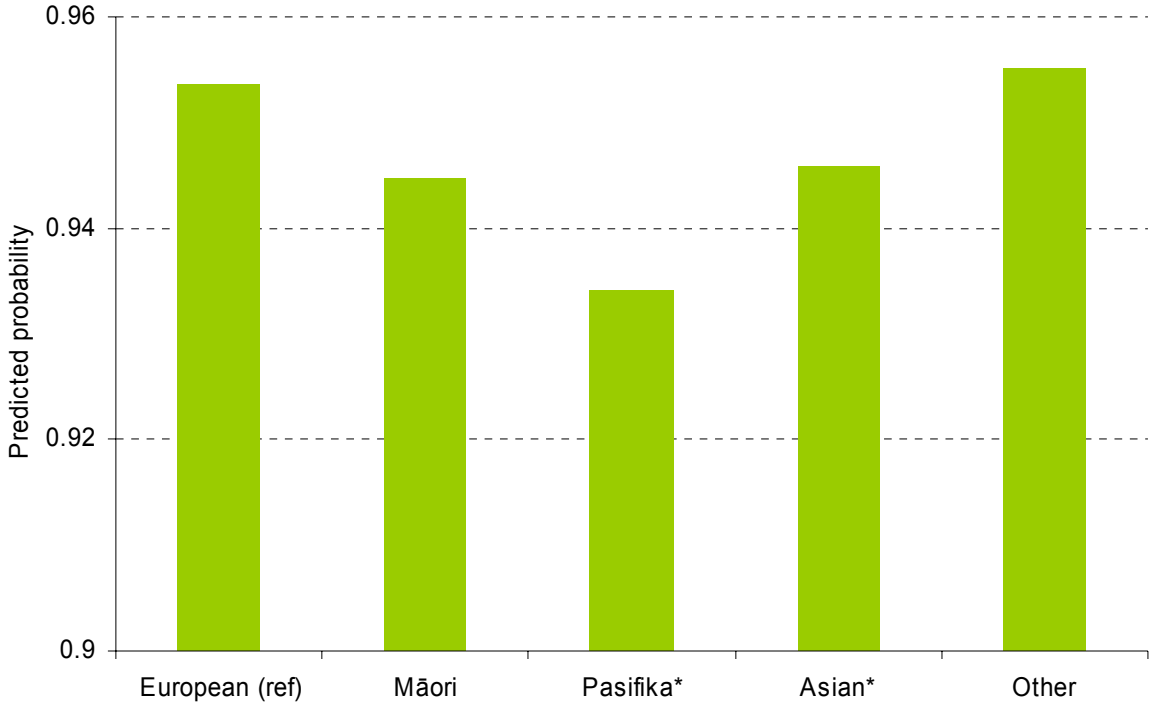
4.2.4 The impact of ethnic group on retention

The ethnic group of a doctoral student also impacted on the likelihood of retention, although the magnitude of the difference was relatively small. Figure 5 presents the predicted probability of retention for a doctoral student in the selected reference group, allowing the ethnic group of the student to vary. Students in the 'Other' ethnic group had the highest rate of retention, followed by Europeans. Although Māori, Pasifika and Asian students exhibited a slightly lower rate of retention than Europeans, the difference was only statistically significant for Pasifika and Asian students.[18]

The RDC component of the PBRF applies a higher funding weighting for the completion of a postgraduate research-based course by a Māori or Pasifika student. This is designed to encourage the completion of a postgraduate research qualification by students in these ethnic groups. As discussed above, the results of the regression analysis indicate that the lower likelihood of retention for Pasifika doctoral students compared with European students is statistically significant, although the difference is relatively small. The regression analysis did not find a statistically significant interaction effect between ethnic group and the presence of the PBRF, which suggests that the likelihood of retention for the various ethnic groups has not changed following the introduction of the PBRF.

However, this finding should be treated with caution as the number of Pasifika students is relatively low - just over 2 percent of doctoral enrolments between 2001 and 2006 were by Pasifika students – which makes it difficult to draw any definitive conclusions. Also, this study is focused on the retention of doctoral students, whereas an analysis of the effectiveness of funding weightings will require an analysis of completions.

Figure 5: Predicted probability of retention by ethnic group



Notes:
1. (ref) identifies the reference category in the logistic regression equation.
2. * indicates whether the category was statistically significant from the reference category at the 5 percent level.
3. The reference group in this analysis was a student aged under 25 in their first year of doctoral study, studying society and culture on a full-time basis at an average New Zealand university in the period prior to the introduction of the PBRF.

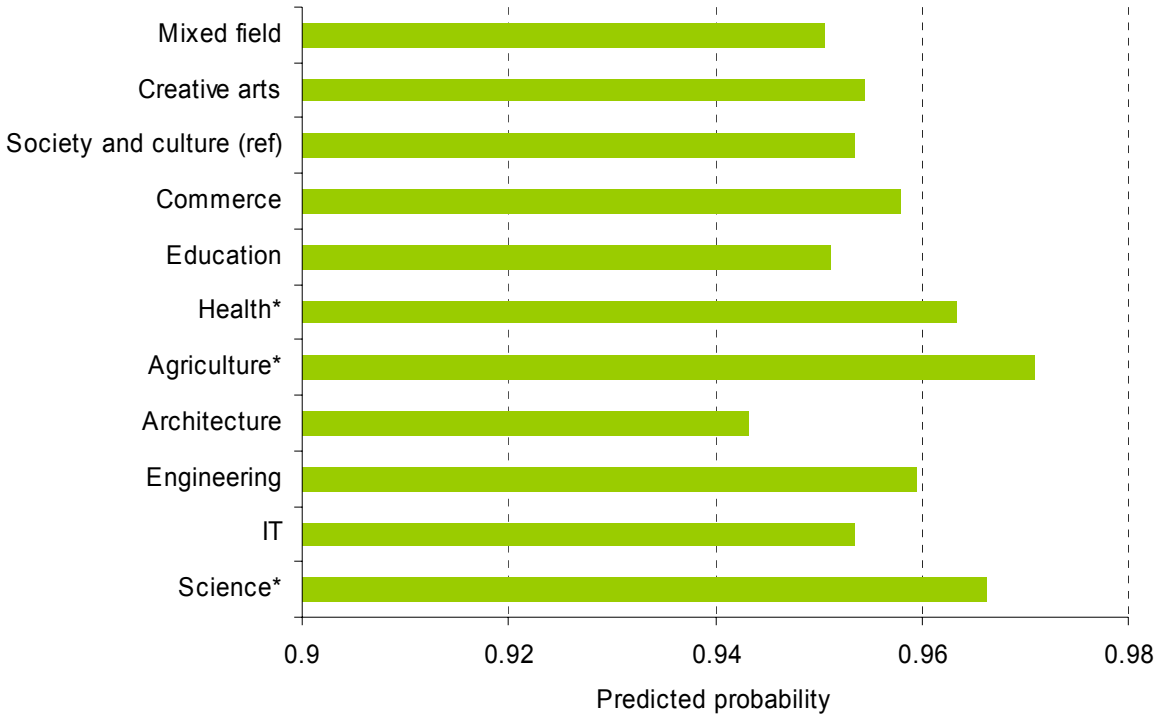
[18] In the case of Pasifika students, this result may partly reflect socioeconomic effects rather than any specific effect of ethnicity on the likelihood of retention.

4.2.5 The impact of field of study on retention

Figure 6 presents the predicted probability of retention for the selected reference group by field of study. Of the subjects taken by doctoral students, science, agriculture and health generally had a higher retention rate than for society and culture, although the magnitude of the difference in the likelihood of retention was small.

The higher retention rate for students in the sciences, agriculture and health mirrors the results of studies by HEFCE (2005) and Martin et al (2001) of completion rates at the doctoral level in the United Kingdom and Australia, respectively.

Figure 6: Predicted probability of retention by field of study

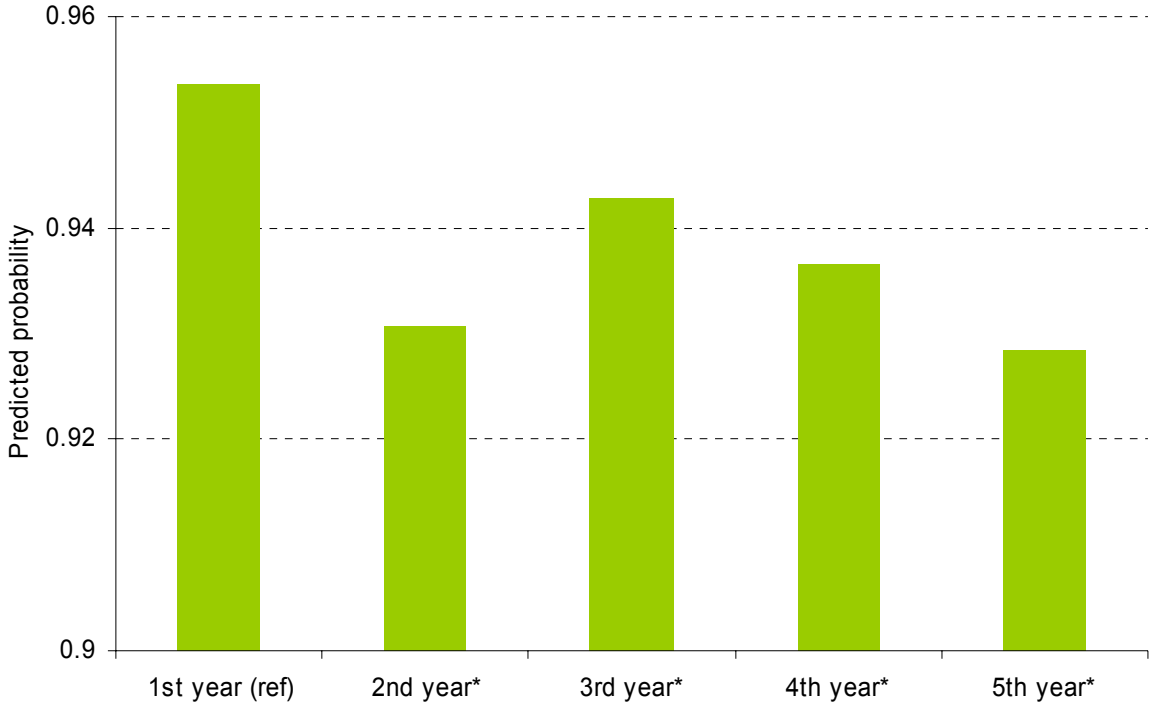


Notes:
1. (ref) identifies the reference category in the logistic regression equation.
2. * indicates whether the category was statistically significant from the reference category at the 5 percent level.
3. The reference group in this analysis was a European student aged under 25 in their first year of doctoral study, studying on a full-time basis at an average New Zealand university in the period prior to the introduction of the PBRF.

4.2.6 The impact of year of study on retention

The regression modelling showed that students in their first year of doctoral studies were the most likely to be retained from one year to the next, with the likelihood of retention generally declining the longer the student has been studying towards their doctorate (see Figure 7). However, the magnitude of this effect is relatively small.

Figure 7: Predicted probability of retention by year of study



Notes:
1. (ref) identifies the reference category in the logistic regression equation.
2. * indicates whether the category was statistically significant from the reference category at the 5 percent level.
3. The reference group in this analysis was a European student aged under 25, studying society and culture on a full-time basis at an average New Zealand university in the period prior to the introduction of the PBRF.

5 Conclusion

The introduction of the PBRF has created significant financial incentives for tertiary institutions to ensure that students who begin doctoral study successfully complete their qualification. In response, the universities have made significant changes to the way they manage postgraduate research. Although it is too early to analyse the impact of these changes on the completion rate of doctoral students, the analysis of the factors that influence the retention of doctoral students has identified some initial trends in retention that may foreshadow the impact of the PBRF on completion.

The findings in this report show that the time period following the introduction of the PBRF has been associated with a small increase in the likelihood of retention of younger full-time doctoral students and a small decrease in the likelihood of retention of older part-time doctoral students, with the trends increasing in impact over time. These trends in retention are likely to at least partly reflect a greater focus on the pastoral care of doctoral students, more rigorous monitoring of progress by doctoral candidates and perhaps different selection priorities, following the introduction of the PBRF.

However, the small decrease in the likelihood of retention for older part-time doctoral students associated with the introduction of the PBRF also coincides with an apparent tightening of the labour market for those with higher tertiary qualifications – illustrated by a fall in the unemployment rate for those with a bachelors or higher degree and a fall in the proportion of starting domestic students who are part-time. Therefore, the tight labour market may also be a factor in this small decrease in the likelihood of retention of this group of students.

The analysis of retention rates showed that the likelihood of retention also varies by ethnic group, with Pasifika and Asian students having a slightly lower likelihood of retention rates than European students. Also, students who studied in the sciences and students in the early stages of their doctoral studies had a higher likelihood of retention. The study also found that the gender and residency status of doctoral students did not have a statistically significant impact on the likelihood of retention.

Further research, such as an examination of the factors associated with the completion of postgraduate research degrees when data becomes available, will more clearly identify whether the PBRF has had the desired effect of raising the completion rate of postgraduate research students.

Appendix A: Summary statistics and regression output

Table 1: Summary statistics for logistic regression dataset

Explanatory variable	Categories	Breakdown of doctoral student population			Retention rate (%)
		All (%) (n = 20,979)	Retained (%) (n = 19,454)	Not retained (%) (n = 1,525)	
RET	Retained	92.7			
	Not retained	7.3			
AGE	<25	12.1	12.6	6.3	96.2
	25-29	23.6	24.1	16.7	94.8
	30-39	31.2	30.9	35.1	91.8
	40-49	21.3	21.0	25.8	91.2
	50+	11.8	11.5	16.1	90.1
GENDER	Male	50.1	50.1	50.5	92.7
	Female	49.9	49.9	49.5	92.8
INTERN	Domestic	89.4	89.3	90.6	92.6
	International	10.6	10.7	9.4	93.5
ETHNIC	European	64.6	64.6	63.7	92.8
	Māori	6.0	5.9	7.1	91.3
	Pasifika	2.1	2.0	2.8	90.5
	Asian	16.5	16.6	15.9	93.0
	Other	10.8	10.9	10.6	92.9
SUBJECT	Science	26.6	27.3	18.6	94.9
	Information technology	4.4	4.3	5.1	91.5
	Engineering	8.9	9.0	7.6	93.8
	Architecture	0.9	0.9	1.3	89.9
	Agriculture	3.2	3.2	2.8	93.7
	Health	11.0	11.1	9.6	93.6
	Education	8.2	7.9	12.1	89.3
	Commerce	7.2	7.1	8.8	91.2
	Society and culture	26.8	26.4	31.1	91.5
	Creative arts	2.5	2.5	2.6	92.4
	Mixed field	0.3	0.3	0.3	91.8
YEARS_STUDY	1	22.3	22.8	17.0	94.5
	2	20.5	20.5	20.7	92.6
	3	18.4	18.6	15.1	94.0
	4	15.4	15.4	15.3	92.8
	5+	23.4	22.7	31.9	90.1
STUDY_LOAD	Full-time	75.6	77.0	57.0	94.5
	Part-time	24.4	23.0	43.0	87.2
PBRF	Pre-PBRF	36.5	36.3	39.1	92.2
	$PBRF_{03}$	19.8	20.0	17.3	93.7
	$PBRF_{04}$	21.2	21.3	19.4	93.3
	$PBRF_{05}$	22.5	22.4	24.1	92.2

Table 2: Logit regression estimates

(Dependent variable: *RET* =1, otherwise 0)

Explanatory variable	Categories	Model 1		Model 2	
		Coefficient	Std error	Coefficient	Std error
AGE	<25	REFERENCE CATEGORY			
	25-29	-0.073	0.126	0.018	0.192
	30-39	-0.398**	0.121	-0.372*	0.179
	40-49	-0.315**	0.127	-0.096	0.191
	50+	-0.390**	0.137	-0.051	0.209
ETHNIC	European	REFERENCE CATEGORY			
	Māori	-0.180	0.110	-0.184	0.110
	Pasifika	-0.383*	0.171	-0.369*	0.171
	Asian	-0.158*	0.079	-0.161*	0.079
	Other	0.032	0.092	0.036	0.092
SUBJECT	Science	0.333**	0.083	0.332**	0.084
	Information technology	0.007	0.132	-0.001	0.132
	Engineering	0.141	0.116	0.142	0.116
	Architecture	-0.229	0.249	-0.211	0.251
	Agriculture	0.474**	0.186	0.484**	0.185
	Health	0.250*	0.101	0.246*	0.102
	Education	-0.046	0.096	-0.053	0.096
	Commerce	0.104	0.108	0.104	0.108
	Society and culture	REFERENCE CATEGORY			
	Creative arts	0.030	0.178	0.021	0.178
	Mixed field	-0.027	0.505	-0.066	0.505
YEARS_STUDY	1	REFERENCE CATEGORY			
	2	-0.422**	0.089	-0.425**	0.089
	3	-0.221*	0.096	-0.220*	0.096
	4	-0.335**	0.096	-0.329**	0.096
	5+	-0.469**	0.085	-0.457**	0.085
STUDY_LOAD	Full-time	REFERENCE CATEGORY			
	Part-time	-1.111**	0.067	-0.908**	0.099
PBRF	Pre-PBRF	REFERENCE CATEGORY			
	$PBRF_{03}$	0.296**	0.078	0.165	0.270
	$PBRF_{04}$	0.195**	0.075	0.604*	0.296
	$PBRF_{05}$	-0.057	0.071	0.726*	0.316
PBRF × *AGE*	$PBRF_{03}$ × (25-29)	-		0.222	0.327
	$PBRF_{03}$ × (30-39)	-		0.380	0.302
	$PBRF_{03}$ × (40-49)	-		0.211	0.315
	$PBRF_{03}$ × (50+)	-		0.046	0.341
	$PBRF_{04}$ × (25-29)	-		-0.264	0.346
	$PBRF_{04}$ × (30-39)	-		-0.051	0.325
	$PBRF_{04}$ × (40-49)	-		-0.491	0.333
	$PBRF_{04}$ × (50+)	-		-0.452	0.357
	$PBRF_{05}$ × (25-29)	-		-0.506	0.360
	$PBRF_{05}$ × (30-39)	-		-0.504	0.340
	$PBRF_{05}$ × (40-49)	-		-0.824*	0.350
	$PBRF_{05}$ × (50+)	-		-1.163**	0.363
PBRF × *STUDY_LOAD*	$PBRF_{03}$ × part-time	-		-0.260	0.160
	$PBRF_{04}$ × part-time	-		-0.377*	0.156
	$PBRF_{05}$ × part-time	-		-0.334*	0.153
CONSTANT		3.626**	0.147	3.416**	0.184
Log likelihood		-5,129		-5,112	
LR $\chi 2$(Prob>$\chi 2$)		<0.0000**		<0.0000**	
Pseudo R2		0.060		0.063	
N		20,956		20,956	

Notes:

1. *, ** represents significant at the 5 percent and 1 percent levels of significance, respectively.
2. Dummy variables for individual institutions were included in the regression models but are not reported here.
3. Robust standard errors that control for the presence of heteroscedasticity are reported in this table.

Appendix B: Regression model using alternative PBRF specification

This appendix presents the results of an alternative specification of the PBRF explanatory variable to test the robustness of the findings in section 4.2.3. In this specification the PBRF is represented by an explanatory variable (*PBRF_ALT*) that takes a value of 1 if the doctoral student was enrolled in study between 2003 and 2005, otherwise 0. This differs from the specification used in equation 1 which had separate categories for each year the PBRF has been in effect. This new specification assumes that the impact of the PBRF does not change over time.

In this alternative specification a variable representing time can be included in the model (*TIME*), which controls for any changes in retention that may be a result of ongoing improvements to the operation of postgraduate research training at the universities not related to the introduction of the PBRF. In equation 2, *TIME* takes a value of 1 in 2001, 2 in 2002 and so on.

The regression equation for the alternative specification took the form:

$$(2)\quad \ln[RET_i/(1-RET_i)] = \beta_1 + \beta_2\, AGE_i + \beta_3\, GENDER_i + \beta_4\, ETHNIC_i + \beta_5\, INTERN_i + \beta_6\, SUBJECT_i + \beta_7\, PART\_TIME_i + \beta_8\, YEARS\_STUDY_i + \beta_9\, PBRF\_ALT_i + \beta_{10} TIME + \beta_{11}\, PROVIDER_i + \mu$$

Where μ is an error term and i = 1 to n observations.

The results of the regression analysis using the alternative PBRF specification are presented in Table 3.

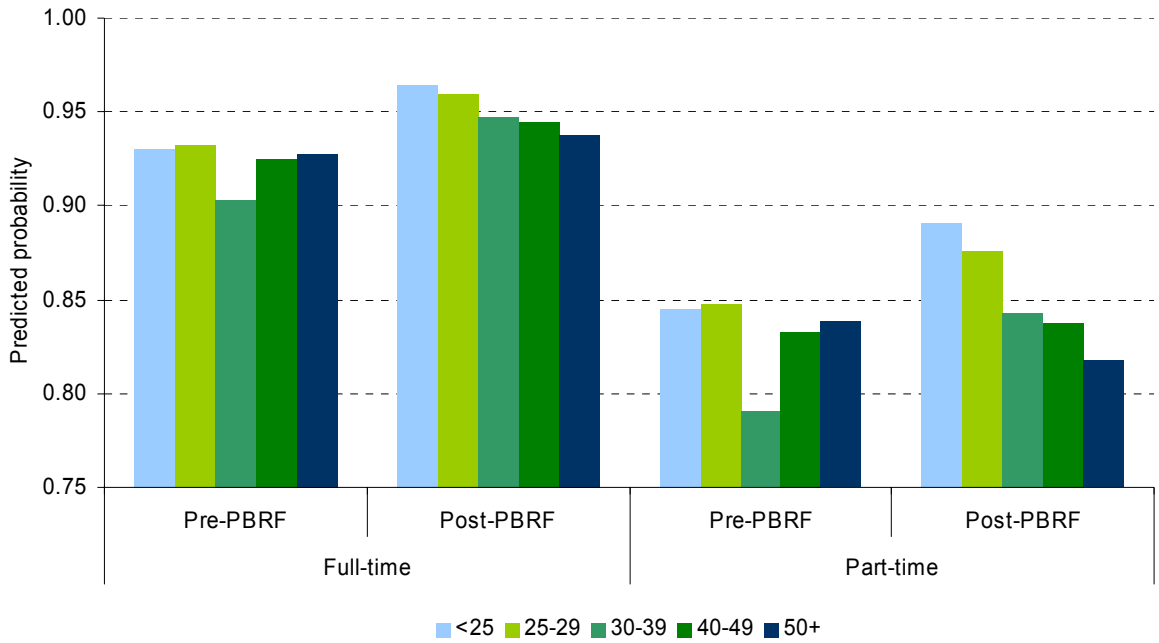
As in the regression analysis in the main section, interaction effects were included in the model between the *PBRF_ALT* and *AGE* and the PBRF_ALT and *STUDY_LOAD*. *GENDER* and *INTERN* were not significant and were removed from the model. The coefficient for the *TIME* variable was statistically significant and had a negative coefficient. This indicates that there was a slight underlying downward trend in the likelihood of retention over time, which may in fact suggest that the PBRF effect illustrated in the original model specification slightly underestimates the full impact of the PBRF.

To illustrate the impact of the PBRF under the alternative specification, the predicted probabilities for the reference group using this new specification are shown in Figure 8. This shows a similar result to the original model specification – ie the introduction of the PBRF has been associated with a slight increase in the likelihood of the retention of younger full-time doctoral students. The predicted probability of retention for full-time students in the under 25 age group increased from 0.93 pre-PBRF to 0.96 post-PBRF.

Figure 8 also shows that prior to the PBRF there was little difference between the predicted probability of retention between the younger and older age groups; however, post-PBRF, older doctoral students had a lower predicted probability of completion than younger students. This is especially noticeable for part-time doctoral students aged over 50, where the predicted probability of retention was lower post-PBRF (0.82) than pre-PBRF (0.84).

Overall, these findings are similar to those presented in section 4.2.3 and confirm that the time period following the introduction of the PBRF has been associated with a slightly higher likelihood of retention for younger full-time students and a slight fall in the likelihood of retention for older part-time students.

Figure 8: Predicted probability of retention by age group, study load and PBRF status (alternative specification)



Note: The reference group in this analysis was a European student aged under 25, studying society and culture on a full-time basis at an average New Zealand university.

Table 3: Logit regression estimates – alternative specification of PBRF variable

(Dependent variable: *RET* =1, otherwise 0)

Explanatory variable	Categories	Coefficient	Std error
AGE	<25	REFERENCE CATEGORY	
	25-29	0.019	0.193
	30-39	-0.370*	0.179
	40-49	-0.094	0.192
	50+	-0.047	0.210
ETHNIC	European	REFERENCE CATEGORY	
	Māori	-0.187	0.110
	Pasifika	-0.374*	0.171
	Asian	-0.163*	0.079
	Other	0.033	0.093
SUBJECT	Science	0.332**	0.084
	Information technology	-0.002	0.131
	Engineering	0.140	0.116
	Architecture	-0.221	0.251
	Agriculture	0.484**	0.185
	Health	0.244*	0.101
	Education	-0.057	0.096
	Commerce	0.105	0.108
	Society and culture	REFERENCE CATEGORY	
	Creative arts	0.017	0.178
	Mixed field	-0.060	0.502
YEARS_STUDY	1	REFERENCE CATEGORY	
	2	-0.421**	0.090
	3	-0.217*	0.096
	4	-0.330**	0.095
	5+	-0.462**	0.084
STUDY_LOAD	Full-time	REFERENCE CATEGORY	
	Part-time	-0.903**	0.104
PBRF_ALT	Pre-PBRF	REFERENCE CATEGORY	
	Post-PBRF	0.710**	0.239
TIME		-0.092*	0.039
PBRF_ALT × AGE	Post-PBRF× (25-29)	-0.169	0.252
	Post-PBRF× (30-39)	-0.052	0.234
	Post-PBRF× (40-49)	-0.371	0.245
	Post-PBRF× (50+)	-0.550*	0.262
PBRF _ALT × STUDY_LOAD	Post-PBRF× part-time	-0.169**	0.252
CONSTANT		3.563**	0.198
Log likelihood		-5,126	
LR $\chi 2$(Prob>$\chi 2$)		<0.0000**	
Pseudo R2		0.061	
N		20,956	

Notes:

1. *, ** represents significant at the 5 percent and 1 percent levels of significance, respectively.
2. Dummy variables for individual institutions were included in the regression models but are not reported here.
3. Robust standard errors that control for the presence of heteroscedasticity are reported in this table.

References

- Coster, G. (2006) Research degree programmes and postgraduate students in Bakker, L., Boston, J., Campbell, L. and Smyth, R. eds. *Evaluating the Performance-Based Research Fund,* pp. 337-374, Wellington: Institute of Policy Studies.

- HEFCE (2005) *PhD research degrees. Entry and retention*, UK: Higher Education Funding Council for England.

- King, G., Tomz, M. & Zheng, L. (1999) *RELOGIT: Rare events logistic regression*, version 1.1, Cambridge, MA: Harvard University, 1 October, http://gking.harvard.edu/.

- King, G. & Zheng, L. (2001) Logistic regression in rare events data, *Political analysis*, 2001:9, pp. 137-163.

- Martin, Y., Maclachlan, M & Karmel, T. (2001) *Postgraduate completion rates*, Occasional paper series, Canberra: Department of Education, Training and Youth Affairs.

- McGinnis, C. (2001) Researching the first year experience: Where to from here? *Higher education research and development*, Vol.20 No.20, pp. 105-114.

- McGivney, V. (2003) *Staying or leaving the course. Non-completion and retention of mature students in further and higher education*, 2nd ed, Leicester: NIACE.

- Ministry of Education (2007) http://educationcounts.edcentre.govt.nz/EducationCounts/indicators/edachievmnt/tspar5.html (accessed 13 September, 2007).

- PBRF Working Group (2002) *Investing in excellence – the report of the PBRF Working Group*, Wellington: Ministry of Education and Transition Tertiary Education Commission.

- Ramanathan, R. (1998) *Introductory Econometrics with Applications,* 4th edition, Florida: Harcourt Brace College Publishers.

- Scott, D. (forthcoming) *Review of TSEC matching. The derivation of unique student identifiers in historical tertiary education data*, Wellington: Ministry of Education.

- Scott, D. & Smart, W. (2005) *What factors make a difference to getting a degree?* Wellington: Ministry of Education.

- Smart, W. (2006) Performance-Based Research Fund: An analysis of the incentives for research degree completion in Bakker, L., Boston, J., Campbell, L. and Smyth, R. eds. *Evaluating the Performance-Based Research Fund,* pp. 159-183, Wellington: Institute of Policy Studies.

- Statacorp (2005) Stata Statistical Software: Release 8.2, College Station, TX: Stata Corporation.

- Tertiary Education Advisory Commission (2001) *Shaping the funding framework*, Wellington: Tertiary Education Commission.

- University of Auckland (2004) *Doctoral Completions Awards*, Auckland: University of Auckland, http://www.postgrad.auckland.ac.nz/postgraddegree/Currentpgstudents/CGSCompAwards.asp (accessed 9 January 2006).

- Williams, C. (1982) *The early experiences of students on Australian Colleges of Advanced Education campuses*, Sydney: University of Sydney.

MINISTRY OF EDUCATION
Te Tāhuhu o te Mātauranga

newzealand.govt.nz