Starpath
A University of Auckland Partnership for Excellence

THE UNIVERSITY OF AUCKLAND
NEW ZEALAND
Te Whare Wānanga o Tāmaki Makaurau

Availability of NCEA Standards: Impact on Success Rate

Starpath Project
The University of Auckland
Epsom Campus, Faculty of Education
Private Bag 92019, Auckland 1142
starpath@auckland.ac.nz
www.starpath.auckland.ac.nz

Bibliographic citation

Turner, T. R., Irving, S. E., Li, M., Yuan, J. (2010) Availability of NCEA Standards: Impact on Success Rate. Auckland: Starpath Project, The University of Auckland.

Disclaimer

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

Other Starpath Project reports may be available from the Starpath website at www.starpath.ac.nz

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Acknowledgements

The authors thank all the schools that provided data for this project.

The authors also thank the members of the Academic Advisory Group for the Starpath Quantitative Team, for their useful advice and input. Special thanks are due to Gavin Brown whose insights led directly to the inception of this project. Thanks are also due to the other members of the Starpath Project who made many useful suggestions and provided a great deal of sound feedback. Margaret Taurere, University of Auckland Equity Advisor, made several very useful suggestions.

1. EXECUTIVE SUMMARY

Research conducted by Starpath (see Turner, 2007) has revealed that Māori and Pacific students attempt, on average, far fewer NCEA Level 3 standards than do their Pākehā and Asian counterparts. This fact is particularly striking in respect of standards from the “Approved List” of subjects. This phenomenon has a serious deleterious impact on the prospects for success of Māori and Pacific students in achieving entrance to university. The current study was initiated to investigate the possibility that part of the reason for the deficit in the number of standards attempted might be a lack of availability of standards.

As a first step in the investigation all New Zealand high schools in deciles 1 to 5 (wherein Starpath’s research interests are focused) were surveyed to determine how many NCEA Level 3 standards were offered by these schools in the 2007 academic year. Requests for data were sent out to schools in early December 2007, with follow-up requests and reminders being sent out in January and February 2008. Of the 226 schools surveyed, 108 responded. Data entry was effected in March 2008, and the data so obtained were merged with the national NCEA Level 3 and higher results for 2007. Data analysis and the writing of the report on the study continued throughout the following year.

The most powerful numeric predictor of academic success turned out to be the number of achievement standards available at the school in question, and this predictor was used throughout the modeling exercise. The analysis consisted in fitting a number of linear and generalised linear models to the data. The most interesting and compelling results came from fitting logistic binomial models, using as predictors the number of achievement standards available, the ethnicity factor, and a factor which captured a particular measure of ability level.

Our major finding is that for Māori and Pacific students in the moderately high ability range the probability of success (at achieving entrance to university according to the NZQA criteria) increases significantly and substantially as the number of available achievement standards increases. In more detail, we initially found that the probability of success for all students is significantly and strongly predicted by the number of available achievement standards. The mechanism by which this predictor influences the success rate appears, however, to be intricate. The number of standards attempted by a student is significantly predicted by the number of standards available to the student, but *not* very strongly: The R^2 value is only a minuscule 0.02.[1]

Moreover the power of the number of available standards to predict success was largely driven by the Pākehā ethnic group (with a lesser contribution from the Māori group). The response curve from the Pacific group was spectacularly flat. That is, the numeric predictor appeared (rather surprisingly) to have no influence at all on the success rate for the Pacific students. However when the results were further subdivided according to an indicator of ability, the numeric predictor had an important effect on Māori and Pacific students in the moderately high ability range. Such students are precisely those for whom we would hope and expect to be able to make a difference.

There is thus persuasive statistical evidence that a focus on efforts to increase the number of achievement standards available to students at a school will have an effective impact in the area where such an impact is most required. The precise policies needed to effect the increase in the number of achievement standards may be difficult to work out — but the message is clear. There is an opportunity to make a difference in the success rate of Māori and Pacific students through judiciously making more achievement standards available to these students.

[1] There is no R^2 value associated with the model, referred to in the previous paragraph, for predicting success from the number of achievement standards available, since the model in question is a *generalised* linear model whose fit is not based on a least squares criterion.

2. INTRODUCTION

Analysis of the NCEA national data sets consisting of results on NCEA standards at Level 3 and higher indicates a strong tendency for Māori and Pacific students to attempt a "smaller quantity" of standards than Pākehā and Asian students. This is true whether "quantity" is measured simply by a count of standards or by the total credit value of the standards attempted. It is also true whether one considers all standards, achievement standards only, or approved list standards only. Bar plots of the quantity of standards attempted by the various ethnic groups (in 2007) are shown in Figure 1 on page 41. Confidence intervals for the means of the quantities are shown in Table 1 on page 40.

It was conjectured that at least part of the reason for the deficit in the number of standards attempted might be a lack of availability of standards. That is, it might be the case that the schools attended predominantly by Māori and Pacific students could be schools that simply are not offering or are not able to offer as many standards to their students as other schools. In order to investigate this conjecture a data-gathering exercise was undertaken. The idea was to acquire information about the "quantity" of standards available at different schools, and then relate this quantity to students' performance in terms of success at achieving university entrance according the NZQA criteria. In what follows we abbreviate this success measure as "NZQA UE".

It must be emphasised at the outset that in this study the concept of "availability" is treated only from a very restricted point of view. That is, if a standard is listed by a school as being available at that school, then it is treated as being available to any student at that school. There are any number of possible complications which cannot be dealt with because the necessary information is impossible to obtain. One such issue is that of prerequisites — if a student (for whatever reason) lacks the prerequisites for the subject containing the standard, then the standard in question is not actually available to him or her. Another factor constraining availability within a school is timetabling which may in effect prevent some standards being available to some students. A related factor is the issue of class size. The standard in question may be available, but in a subject which has only one class where the number in the class is limited to, say, 20. Students who are tardy in making their subject choices for the year may find themselves excluded from that class, so that the standard in question is again not actually available to them. A further important issue is that students may be guided or even forced to make particular subject choices whereby the availability of certain standards is effectively removed for them. Despite these limitations the study as actually conducted has yielded interesting results.

Throughout this report we follow the usual convention that, for brevity, the word “significant” is used to mean *statistically* significant (at the 0.05 significance level). Likewise “significantly different” is used to mean *statistically* significantly different.

3. LITERATURE REVIEW

In this report we undertake the study of the impact of the number of NCEA standards, available to students at a given school, upon those students' chances of achieving UE according to the NZQA criteria. There appears to be little in the literature which is germane to this topic. There is a small amount of literature which deals with the provision of course offerings from which students make choices that influence future choices and which ultimately affect further education and career options. These investigations are somewhat related to the concerns of the current study.

Finn and colleagues (Finn 1997; 1999; Finn, Gerber, and Wang 2001; 2002) considered the issue of course offerings and practices that limit or increase the course work (of various types) taken by particular groups of students in the USA, and have noted the deleterious effects of tracking or streaming. Finn (1997) asserted that there were unambiguous findings about the relationship of course-taking to academic achievement.

Advanced courses and "gatekeeper[2]" courses were not equally available in all schools or to all students. As the proportion of low-income and minority students in schools increased, the relative proportion of college-preparatory and advanced course sections decreased. Another study (Finn, 1999) found that breadth and depth of course offerings were consistently lacking in schools in small and rural communities and that there were also problems in schools with low socioeconomic status. This study also found that African American and Hispanic students tended to take fewer mathematics, science, and foreign language courses which directly parallels the New Zealand experience in respect of Māori and Pacific students.

In respect of mathematics (Finn *et al.*, 2001 and Finn *et al.*, 2002), the authors examined course offerings and course taking patterns, and found differences in course offerings with respect to rural and urban schools, and with respect to smaller and larger schools, but not with respect to high- and low-poverty schools. However, they did find a significant association of course taking with school poverty level. Tracking or streaming was found to have a significant (negative) impact on course taking patterns. Course offering and

[2] A "gatekeeper course" is one which is a prerequisite to a particular academic pathway or program. Such courses are usually designed to be academically demanding and thus have the effect of precluding many students from pursuing the pathway in question.

graduation requirements had different impacts on course taking depending on students' racial/ethnic identity and the track to which they had been assigned.

Lee and Bryk (1988) argued that placement in a specific track and the consequent course of study were major mediating factors in linking student backgrounds with academic achievement, and that placement in non-academic tracks was associated with lower levels of achievement.[3] Amongst other things, Lee and Bryk (1989) found that high levels of achievement were related to the "academic emphasis" of a school. This closely parallels conclusions drawn in the current study.

Lee and colleagues noted the paradoxical effect of "increased choice" actually reducing the likelihood of equitable outcomes. For example, Lee (1993) showed that a wide latitude of student choice of courses in high school magnified the social stratification of educational outcomes. This socially undesirable consequence results from two well-documented relationships: (a) Following a more demanding set of academic courses in high school is strongly and positively associated with higher academic achievement, and (b) less advantaged students are considerably less likely to select such a demanding course of study than are their more academically and socially advantaged counterparts.

In analyzing NAEP[4] data from 123 schools, Lee, Croninger and Smith (1997) found that students learn more in schools which offer a narrower curriculum of academic courses. In Israel, however, Ayalon and Gamoran (2000) found that greater diversity of course offerings was related to student selection of higher level courses and to higher average achievement and greater equality of achievement.

A number of studies on students' course choices within the NCEA system has been conducted in New Zealand over the past eight years. The focus of these studies has been on how schools have used the flexibility of the NCEA system to respond to the learning needs of their students and create different educational pathways (Hipkins & Vaughan, 2002; Hipkins, Vaughan, Beals & Ferral, 2004; Hipkins, Vaughan, Beals, Ferral & Gardiner, 2005), and on the effects of students' course choices on their motivation and achievement (Meyer, McClure, Walkey, McKenzie & Weir, 2006; Meyer, Weir, McClure, Walkey & McKenzie, 2007). A study

[3] Schools might argue that students with lower levels of achievement *should* be placed in non-academic tracks. This ignores the fact that some low achievers might have a high level of academic ability and simply need the appropriate encouragement or circumstances in order to demonstrate this ability.

[4] National Assessment of Educational Progress; an American initiative which attempts to provide a "nationally representative and continuing assessment of what America's students know and can do in various subject areas." See for instance http://nces.ed.gov/nationsreportcard/about/ .

undertaken by a Starpath team (Madjar, McKinley, Jensen & van der Merwe, 2009) showed that students are not always the ones making the choices, that schools have a strong mediating role in what subjects are available to which students, how informed students and their parents are about course choices, and how early choices and performance can shape students' course options at more advanced NCEA levels. None of these studies tested the relationship between course (or standards) availability and UE success rates.

In summary, there is only a moderate amount of discussion, in the literature, of the relationship between course offerings/taking and academic achievement, and of the consequences for important sub-groups within schools. Overall, there appears to be little consensus on the conclusions to be drawn from the investigations that have been undertaken, and policy implications remain unclear.

4. COLLECTING THE DATA

For the purpose of obtaining data on the availability of standards, requests were sent to all schools in New Zealand which had decile rankings in the 1 to 5 (inclusive) range and which offered at least one Year 13 subject. Attention was restricted to deciles 1 to 5 since it is schools in these deciles which come within the purview of the Starpath Project. Schools were asked to provide us with a list of all NCEA standards, at Level 3 and higher, which were available to their students in the 2007 academic year.

Due to the timing of the request some schools sent out listings of standards that were available in 2008 rather than in 2007. Apparently in many instances the 2007 listings were no longer readily available to the schools at the time at which they attended to our request. The 2008 listings were used on the assumption that the 2008 offerings would be approximately equal to those of 2007.

There were 226 schools in the population of interest and of these 108 provided responses. The data were summarised in terms of the total number of (Level 3 and higher) standards offered by each school, and the total number of credits at Level 3 and higher. These values were also subdivided according to whether the standards were unit or achievement standards, and whether they were from the Approved List of subjects. The resulting data were then merged with the appropriate data from the NCEA 2007 results, on the basis of the "provider code" of the schools in the data set.

We also extracted from the NCEA 2007 results those records corresponding to the 118 schools to which requests were sent but which did *not* respond. This was done for the purpose of comparing, as much as possible, the responding schools with the non-responding ones with a view to detecting any systematic differences or sources of bias. This comparison is discussed in Appendix 4 (page 39).

5. DATA ANALYSIS

Questions to be asked of the data involve the “quantity” of standards available at schools. The notion of “quantity” could be specified in various ways, such as total number of standards available or total number of credits available. The standards under consideration could also be restricted to achievement standards or standards from the Approved List. Fitting models using the various options led to the conclusion that the number of achievement standards available was the best or effectively equal-best predictor in all instances. Henceforward we focus on number of achievement standards as the measure of “quantity” to be used.

Specific questions included:

1. Does the number of achievement standards available at a school have an impact upon the number of such standards attempted by students?
2. Is there an indication of a relationship between the ethnicity of a student and the number of achievement standards available at the school attended by that student?
3. Does the number of achievement standards available at a school affect the probability of its students achieving NZQA UE?
4. Does the number of achievement standards available at a school affect the “GPA” score (see subsection 5.4) of its students on NCEA standards?
5. If the number of achievement standards has an effect, how big is the effect, or how important is it, in practical terms?
6. What is the relationship of the effect (if there is one) of the number of achievement standards to ethnicity, and does the effect change with ethnicity?
7. How do other factors such as decile and school size interact with the number of achievement standards available in respect of predicting student outcomes?

Attention was restricted to students whose year of study was Year 13 or higher. The reason is that we are interested in students’ success at achieving university entrance according to the NZQA criteria. The concept of “success” here is really meaningful only if students are actually *aiming* at achieving NZQA UE (in the year under consideration, i.e. 2007). There is of course no data available with respect to the students’ intent, but it is reasonable to assume that if a student’s year of study is below 13 then the student is not (yet) aiming at NZQA UE. Likewise if the year of study is 13 or higher then achieving NZQA UE is at least a reasonable expectation. It is definitely possible for brighter students to have already achieved NZQA UE at Year 12,[5] so restricting to Year 13 and above is not completely appropriate. Nevertheless, some sort of restriction must be used, and the Year 13 restriction appears to be the only practical one.

5.1 Number of Standards Attempted vs. Number of Standards Available

The conjecture with which this project started was that part of the reason that Māori and Pacific students attempt fewer Level 3 standards (especially achievement standards) is that fewer standards are available at the schools which they generally attend. This conjecture is based on the expectation that the number of standards attempted should increase with the number of standards available. Figure 2 on page 42 shows a plot of the number of achievement standards attempted versus the number of such standards available at the students' schools. The correlation between the two variates is about 0.1535 (95% confidence interval [0.1309,0.1760]) — significantly positive, but not large. Thus the expectation appears to be true to some extent, but the impact is neither large nor visually obvious.

The solid line which appears in Figure 2 is the line of least squares fit. As one might expect from the appearance of this plot, the R^2 value for the fit is minuscule: 0.02. The dotted line which appears in this plot is the "$y = x$" line which shows that 6 students appear to have attempted *more* achievement standards than are actually available at their schools. One possible explanation for this is that these students attempted these standards via the Correspondence School, or at another nearby school. Another possible explanation is that (as discussed in section 2) the availability data from many schools was for 2008 whereas the students' results are for 2007.

Fitting a linear model to predict the number of achievement standards taken from the number available, and including ethnicity as a predictor, reveals that the number of available standards and ethnicity are both highly significant predictors, as is their interaction. The slopes of the regression lines were significantly positive for all groups but the Pacific group. The *p*-values were 0, 0.004, $1.77{\times}10^{-5}$, and 0.003 for the Pākehā, Māori, Asian and "Other"

[5] According the NZQA web page there were 308 such students in all high schools, New Zealand wide. There were also seven Year 11 students who achieved NZQA UE.

groups respectively. The *p*-value for the Pacific group was 0.84. The point estimate of the slope for the Pacific group was (just barely) negative: −0.0015.

A plot of the data, subdivided by ethnic group, with fitted lines superimposed is shown in Figure 3 on page 43. Adding ethnicity to the model provides a substantial increase to the R^2 value, as compared with the model in which the number of standards available is the only predictor. However, as one might expect from Figure 3, the value remains very small—it is only 0.11.

5.2 Number of Standards Available by Ethnicity

If the lack of available standards were the explanation (or a major part of it) for the fact that the Māori and Pacific ethnic groups attempt fewer standards than the Pākehā and Asian groups, then we would expect there to be differences in the number of standards available to students classified by ethnic group. A boxplot of the result (for the number of *achievement* standards) is shown in Figure 4 on page 44. This plot does indeed appear to indicate deficiencies in the availability of achievement standards for the Māori and Pacific groups, but these deficiencies must be described as moderate.

Running a one-way analysis of variance indeed shows that the mean number of achievement standards available is significantly lower for Māori students, as compared with Pākehā (mean difference 9.4, with a standard error of 1; *p*-value to all intents and purposes equal to 0). However for Pacific students there is no significant difference from Pākehā (*p*-value = 0.527).

It is important to note that the boxplots in Figure 4 indicate that the *median* number of available achievement standards is significantly[6] lower for Pacific students than it is for Pākehā students, despite there being no evidence of a difference in *means*. We remark that the fact that the means and medians are substantially different is an indication that the data are likely to be non-Gaussian. Nevertheless the analysis of variance is valid since the sample sizes involved are large. (The minimum size, for the “Other” ethnic group, is 197; the remaining sample sizes are well over 1000.) Whether mean or median is a more appropriate measure in this context is a separate question which does not appear to have a clear answer.

[6] As a rough approximation at least, the medians of two samples are significantly different at the 0.05 significance level if the notches of boxplots of the two samples do not overlap.

Asian and “Other” students have significantly *more* achievement standards available to them than do Pākehā students as measured by either mean or median. It is interesting to note the large number of lower-tail outliers for the Asian group. What the presence of these outliers is saying is that the bulk of Asian students attend schools where a fairly large number of achievement standards are available. However there is a small number of Asian students who are noticeably “different” in respect of the sort of school that they attend. Any explanations of this phenomenon are necessarily speculative.

The phenomenon may reflect the fact that many Asians are relatively recent immigrants to New Zealand, who can and do choose where to settle and who use the nature of the schools available to their children as one of the criteria for making this choice. A fact that must also be considered is that “Asian” is a very broad categorisation, including Korean, Vietnamese, Indian, Chinese, Indonesian, and many other ethnic or national groupings. The outliers could simply result from the heterogeneity of the Asian ethnic group. Investigating the cause of these lower-tail outliers could be a topic for further research.

5.3 School-by-school Analysis: Success Fraction

We have seen in section 5.1 that the number of achievement standards available has only a small impact on the number of such standards attempted. There may be other mechanisms by which the number of achievement standards available affects student outcomes. To explore this possibility we first looked at the data on a school-by-school basis, and investigated the effect upon the “success fraction”. For simplicity we defined the success fraction in terms of the available data, which is to say the national NCEA Level 3 results in which all students who have attempted at least one Level 3 standard are included.

The success ratio for a school is then the ratio of the number of students at that school achieving NZQA UE to the number of Year 13 (or higher) students of that school appearing in the national NCEA data set. It might be argued that the success ratio would be better defined in terms of the number of “participating students”, i.e. the number of students currently attempting sufficiently many Level 3 standards in order that it be possible for them to achieve NZQA UE. However this is not feasible due to inherent difficulties with the data. It is not practical to determine who is or is not a participating student and information on this issue is not readily available.

In the future it might be possible to proceed on a "participation" basis since in 2008 NZQA began publishing figures on UE success rate on this basis as well as on a school roll basis. Thus it might be possible to obtain the required information from NZQA. However for the year under study this is not feasible. Basing success rate on "participation", or using a minimum larger than "at least one Level 3 standard" would very likely have some impact on the results. In section 5.6, in the context of student-by-student analysis, we consider restricting the data set to students who have attempted at least 30 credits worth of Level 3 standards from the Approved List. This had the effect of making the results somewhat less spectacular, although the fundamental conclusions remained unchanged.

Finally, we remark that we were interested in success at achieving UE irrespective of whether a lack of success was due to attempting too few standards or failing too many standards.

A plot of the success fraction of students at the schools in the study, versus the total number of achievement standards available at the schools, is shown in Figure 5 on page 45. The plot shows a reasonably strong and positive "linear" relationship between the number of standards available and the success fraction. For the main body of the "data cloud" a linear relationship seems eminently plausible. There are however two satellite clouds of outliers (the points above the success rate value of 0.6 and those below the value of 0.05).

The upper cloud corresponds to schools which appear to be "punching above their weight" in terms of success rate. These schools merit further investigation to determine just what it is that they are doing to make them so successful. The lower cloud corresponds to schools which achieved a success rate of 0. These data points also merit further investigation. Such investigation, however, was not part of the study under current discussion.

The fitted line shown is the result of fitting a weighted linear model to the data with the clouds of outliers, discussed above, excluded. The weights used were the reciprocals of estimated variances of the response values (i.e. of the success fractions). The R^2 value for the (weighted) fit was 0.24. With the outliers left in, the corresponding R^2 value was 0.12.

5.4 School-by-school Analysis: NCEA GPA

Another measure by which student outcomes could be assessed is a form of grade point average (GPA) calculated from the students' NCEA results. The form of GPA that is used in this report is the same as that discussed in Shulruf, Hattie, and Tumen (2007) and Turner, Li, and Yuan (2010). That is, it is calculated on the basis of *Approved List* standards only and the scoring system used was 0 points for "not achieved", 2 points for "achieved", 3 points for "merit", and 4 points for "excellence". The GPA is formed as a weighted average of each student's scores, the weights being the number of credits associated with the standards. Clearly a GPA is not meaningful unless some minimum number of credits has been attempted. We have taken the "minimum number" to be 30 credits; students who attempted fewer than 30 credits (in standards from the Approved List) were assigned a "missing value" as their GPA.

Figure 6 on page 44 shows a plot of the mean NCEA GPA for schools versus the number of achievement standards. A "linear trend" is discernible in this plot, but the noise level (i.e. variability about the trend line) is very high. Four "obvious" outliers can be seen in Figure 6. It is interesting to note that one of the three "high" outliers in Figure 6 is the same school as one of the "*low*" outliers which appear in Figure 5 on page 45. One of the other high outliers coincides with a "high" outlier in Figure 5, and the third high outlier does not appear as an outlier in that figure. The one "low" outlier in Figure 6 coincides with a "low" outlier in Figure 5. These outliers all merit further investigation, but such investigation is beyond the scope of this report.

The fitted line shown is the result of fitting a linear model to the data with the above mentioned outliers excluded. The R^2 value for the fit was 0.08 irrespective of whether the outliers were excluded (i.e. excluding the outliers caused no change to the R^2 value to two decimal places).

5.5 Student-by-student Analysis: Binary Success Response

We now turn to a student by student, rather than school-by-school analysis. Looking at "success" (at achieving NZQA UE) on a student-by-student basis means that we are looking at a binary variable (1 for success, 0 for failure). It is not feasible to display such data graphically in a useful manner. We therefore proceed directly to model fitting. The appropriate model is a binomial logistic generalised linear model. The response was success at achieving NZQA UE. The number of achievement standards (available at the school attended by the student in question) was again the best predictor, yielding the smallest residual deviance (analogous to the residual sum of squares in an ordinary linear model).

Once such a model is fitted, the fitted response (estimated *probability* of success) may be plotted against the predictor. Such a plot is shown in Figure 7 on page 47. Also shown is an approximate 95% confidence envelope for the mean success probability. This figure would appear to indicate a substantial influence by the number of achievement standards available upon the success probability. This is borne out by the fitted model: The *p*-value for the predictor is effectively 0.

To assess the impact of the predictor on the success probability, the fitted value when the predictor is equal to the sample mean[7] (109.4) can be compared with that when the predictor equals the sample maximum (166). These fitted values are 0.3971 and 0.4936, respectively (to four decimal places). This means that the increase in the success rate, when the predictor increases from its mean to its maximum, is about 0.1 or 10 percentage points. As a percentage of the starting value (the success rate at the mean of the predictor) this is about (10⁄40)×100 = 25%. The standard error for the increase or change in the fitted value is not available on an analytic basis. It can however be obtained via Monte Carlo methods (parametric bootstrapping). The resulting 95% confidence interval for the change in success probability is [0.07, 0.12]. That is, we are pretty sure that the change is at least 0.07 (or 7%) and at most 0.12 (or 12%).

5.6 Student-by-student Analysis: Binary Success Response with Ethnicity Included as a Predictor

The results of the foregoing analysis are not in themselves startling. The number of achievement standards available is a measure of the school's academic capacity and one might legitimately expect that increasing this capacity would lead, up to a point at least, to an increasing success rate. A vital component is however missing. Ethnicity is likely to be a significant predictor of success probability — and indeed statistical tests confirm this. But more to the point, there is a significant *interaction* between ethnicity and the effect of the number of achievement standards available (*p*-value $\approx$ 0.009).

The nature of this interaction is depicted in Figure 8 on page 48. We see that the success probability increases much more rapidly for the Pākehā ethnic group than for the others, and that for the Pacific group the estimated success rate curve is effectively a flat (horizontal)

[7] This is the mean over students, or equivalently the mean over schools weighted by the number of students at the school.

line. This means that for the Pacific group the success rate appears to be uninfluenced by the number of achievement standards. The estimate of the coefficient of the predictor is in fact nowhere near being significant; the *p*-value for this coefficient (compared against a zero baseline) is about 0.96.

It should be noted that, even though the fitted curves are not as “flat”, the predictor is not significant for the Asian or “Other” ethnic groups either — *p*-values of 0.15 and 0.19 respectively. In contrast, the *p*-value for the predictor is effectively 0 in the case of the Pākehā group. For the Māori group the *p*-value is 0.02. It is worth emphasising that that Figure 8 displays a striking difference between the Māori and Pacific ethnic groups, demonstrating that it is unwise to combine these two groups in a single category as is so often done.

The surprising nature of the interaction between the numeric predictor and the ethnicity factor calls for further investigation. One approach is to further subdivide the students according to their “academic potential”. We have no prior data on such “academic potential”. Instead we used the students’ NCEA GPA score as a surrogate measure of this potential, and subdivided the students into groups of “lo”, “midlo”, “midhi”, and “hi” potential according to the quartiles of this score. The corresponding lattice plot is shown in Figure 9 on page 49.

The results of the fit are fascinating, and we believe, crucially important. The coefficient of the numeric predictor is significantly different from 0 only in two of the 20 “ethnicity by GPA category” cells. These two cells are “(Māori, midhi)” and “(Pacific, midhi)”. The *p*-values for the coefficients of the numeric predictor, corresponding to these two cells, are both 0.001 (to three decimal places). The *p*-values for these coefficients corresponding to the other 18 range from 0.116 to 0.992. The implications of this phenomenon are explored in section 6 of this report.

It is of interest here to examine the actual success rate of students in the various (twenty) categories depicted in Figure 9. These success rates are shown by the bar plots in Figure 10 on page 50. We notice that for the “lo” and “midlo” levels of NCEA GPA the pattern is quite similar amongst the different ethnic groups. The success rate of the Māori and Pacific groups does not seem to increase as rapidly, when one moves up the GPA rankings, as it does for the Pākehā and Asian groups. In fact the success rate of the Pacific group changes very little among the top three GPA categories.

The foregoing analysis, depending as it does upon NCEA GPA scores, was restricted to that subgroup of students for whom a "non-missing" NCEA GPA score is assigned, that is, to students who attempted *at least* 30 credits' worth of Approved List standards. In view of this, it seemed advisable to reanalyse the data, restricting the analysis to the same set of students as were included in the earlier analysis, but without incorporating the GPA categories.

The result was somewhat less spectacular—the response curve for the Pākehā group was less steep than before and that for the Pacific group was less "obviously horizontal". For the Pākehā group the predictor (i.e. number of achievement standards available) was highly significant: p-value equal to 2.4×10^{-5}. For all of the other groups, including the Māori group, it was not significant. (Interestingly the predictor became "border-line significant" for the Pacific group — p-value = 0.0536 — when the data set was restricted.)

When the success probability was modelled only by the numeric predictor (no ethnic group factor included) the result for the restricted group of students was very similar to the result obtained when all students were included — except of course for a "level shift". The restricted group had a much higher intercept term. The coefficient of the numeric predictor differed very little between the two fits.

We also assessed the practical importance of the influence of the number of available achievement standards on the success rate. We did this by calculating the expected change in success probability when the predictor changes from its mean value to its maximum value. (The mean varied according to ethnic group; the maximum was the same — 166 — for all five ethnic groups.) Again, Monte Carlo methods were used to obtain standard errors and hence approximate 95% confidence intervals for the increase in success probability. The calculations for the entire data set (i.e. not restricting to students with a non-missing GPA) yielded the results shown in Table 2 on page 40.

The confidence intervals in Table 2 indicate that there is evidence that the expected change is greater than zero in the case of Pākehā students. There is also evidence, albeit minimal, that the expected change is greater than zero in the case of Māori students. For Pacific, Asian, and "Other" students there is no evidence that the expected change is different from 0.

For the restricted data set the results were as shown in Table 3 on page 40. The expected change is significantly different from 0 only for the Pākehā group. Even though the point estimate of the change for the Pacific group is by far the largest in percentage terms, this change is not significantly different from 0. The confidence interval for the overall change in success probability (with ethnicity ignored) is [0.06, 0.10].

5.7 Direction of Causation

In the previous two sections it has been tacitly assumed that there is a causative relationship between the number of achievement standards available and the success rate. That is, it has been implied that increasing the number of achievement standards available leads to or causes an increase in the success rate. This assumption merits some discussion. It is of course conceivable that the direction of causation is in fact in opposite to that which we have been assuming. That is, it may be that increasing success rate "causes" an increase in the number of achievement standards available in that if the school population includes a large fraction of academically bright students, who will tend to succeed, then the school authorities will take steps to make more achievement standards available to satisfy the needs of these students.

However, if this were the case, then one would expect that *given* the potential level of a student, the number of achievement standards available would make no difference to the student's probability of success. For example, consider student "A" who has high academic potential but happens to attend a school where the bulk of the students show low potential, hence causing the school to offer few achievement standards. At the same time, consider student "B" endowed with similarly high academic potential but who attends a school in which the bulk of the students also show high potential, hence causing the school to offer many achievement standards. Under the scenario that high potential "causes" an increase in the number of achievement standards it would appear that student "A" would have no worse chance of achieving NZQA UE than student "B".

We had no data on the academic potential or ability of the students in the study, but have in section 5.6 used NCEA GPA as a surrogate for such data. The analysis in that section indicates that for some ethnic groups there is still predictive power in the number of achievement standards available even when the quartiles of the NCEA GPA are conditioned on. To clarify the issue, if ethnicity is left out, and the success rate modeled only in terms of the number of achievement standards available and the quartiles of NCEA GPA, then in the "midhi" and "hi" quartiles the success rate responds strongly to the number of achievement standards available. In the "lo" and "midlo" quartiles the numeric predictor is non-significant (marginally so for "midlo" — p-value = 0.06). These results are illustrated in Figure 11 on page 51.

Likewise, if a model is fitted using both NCEA GPA and the number of achievement standards available as continuous predictors, the latter predictor remains highly significant; the p-value is 3.9×10^{-14} (effectively 0). Although short of being "proof positive" that causation proceeds in the direction that we are claiming, nevertheless, it is fairly convincing evidence that causation is in the direction claimed.

It should of course be noted that it is possible that neither direction of causation is an appropriate description of the actual state of affairs. In other words, it is possible that both the success rate and the number of achievement standards available are driven by a third variable, a "hidden covariate". This latter possibility is however too vague to be effectively explored. The direct causation mechanisms discussed above are intuitively plausible, and any obvious related covariates (e.g. school size and decile) have been dealt with.

5.8 Student-by-student Analysis: NCEA GPA Response

When NCEA GPA is plotted against the number of achievement standards available, on a student-by-student basis, the result initially looks completely random. (See Figure 12 on page 52.) Despite appearances, however, the number of achievement standards does have (some) power to predict NCEA GPA. The p-value for this predictor in a one variable regression model is about 7×10^{-8}. The R^2 value for this fit is of course very small: 0.006. The fitted line shown in Figure 12 is essentially the same as that shown in Figure 6 on page 46.

We also undertook an analysis of the fit with ethnic group included as a predictor. In this fit, both ethnicity and its interaction with the number of achievement standards are highly significant. A plot of the data subdivided by ethnic group, with fitted regression lines superimposed, is shown in Figure 13 on page 53. The R^2 value for this fit is tiny: 0.03. (It is of course larger than for the model not including ethnicity, but is still small.)

The slopes of the regression lines for the Māori and "Other" groups are not significantly different from 0. The slope for the Pacific group is significantly (p-value = 0.03) *negative*.

5.9 The Impact of School Decile

School decile forms a rough measure of the average socio-economic status of the students at that school. It might be expected to have an impact upon the courses and subjects emphasised by schools, in particular upon the number of achievement standards available. Figure 14 on page 54 shows a plot of the number of achievement standards available versus school decile for the schools in this study. The relationship between the two variables is roughly linear.[8] The number of achievement standards available seems to increase by about 10 (on average) for each increment in decile level.

The superimposed lines are fitted regression lines that might be used to predict the number of available achievement standards from decile. The upper (green) line is the "student level" prediction. That is, choose a student at random, determine the decile of the school attended by the student, and from that predict the number of standards available to the student. The lower (red) line is the "school level" prediction. That is, choose a school at random, determine its decile and from that predict the number of standards available. In effect the upper line is the regression of the number of standards on decile *weighted* by the number of students (in Year 13 or higher) at the schools. The lower line is the un-weighted regression.

The substantial gap between the two fitted lines (school level and student level prediction) in Figure 14 is indicative of a *size* effect with respect to the number of (achievement) standards available. The student level prediction gives greater weighting to larger schools, which has the effect of increasing the average number of standards. This size effect is further explored in section 5.10.

Given that the school decile predicts the number of standards available, which predicts the UE success rate, it was of interest to see if the number of standards would add any predictive power over and above that which is provided by the decile. To this end we fitted an appropriate model for success rate which involved the three predictors: the number of achievement standards, school decile, and ethnicity. We then tested this model against the simpler model involving just the two predictors: school decile and ethnicity. The *p*-value for the test was 4.4×10^{-7}, i.e. the more elaborate model provided a highly significant improvement in predictive power. Thus the number of achievement standards provides predictive power for success over and above that provided by the school decile.

[8] Although a goodness of fit test shows that the relationship is not, strictly speaking, linear, this is a technicality which is of no immediate concern.

A lattice plot of the resulting fitted success probabilities is shown in Figure 15 on page 55. The appearance of this plot is fascinating; the estimated curves appear to be horizontal for the Māori group and *downward* sloping for the Pacific group. Significance tests obtained from the fitted model confirm this visual impression, i.e. the coefficient of "number of achievement" standards is not significantly different from 0 for the Māori group and is significantly negative for the Pacific group. For the Pākehā, Asian and "Other" groups these coefficients are not significantly different from 0 although the point estimates are positive (as the appearance of the graphs indicates).

5.10 The Impact of School Size

In order to deal with "school size" we needed a measure of this quantity. The most convenient measure, readily available from the data sets with which we worked, was the number of students in Year 13 or higher who attempted at least one Level 3 (or higher) NCEA standard and hence appeared in the national NCEA data set. Using this measure of "school size" we plotted the number of achievement standards available against school size as shown in Figure 16 on page 56.

This figure indicates that there is a strong, although complicated, relationship between school size and the number of achievement standards available at the school. The number of standards increases non-linearly with increasing number of students. The fitted line shown is from a cubic fit; this fit is not completely satisfactory on a couple of grounds, but we were unable to come up with a better one. Given that there is a significant, albeit intricate, relationship between school size and the number of achievement standards offered it would appear that the impact of school size upon success rate should be further explored.

In particular we needed to consider the possibility that school size is the relevant predictor, and that the number of available achievement standards is simply a surrogate for school size. To investigate this we fitted a logistic binomial model, with the (binary) response being success at achieving NZQA UE, and the predictors being school size (modelled in terms of a cubic polynomial) and ethnicity, with appropriate interactions. We also fitted a model which included the foregoing effects and in addition the number of achievement standards and its interaction with ethnicity. The more elaborate model was significant when tested against the reduced model (p-value = 0.0008), indicating that the number of achievement standards is *not* merely a surrogate for school size.

On the other hand, testing the more elaborate model against a reduced model that involved only the number of achievement standards and ethnicity yielded a *p*-value of 9.8×10^{-9}. Hence school size adds predictive power over and above that provided by the number of achievement standards. The pattern of dependence of success probability upon school size is shown in Figure 17 on page 57. The cubic nature of the way the size effect was modelled revealed some interesting structures in the relationship between school size and students' success probability.

All of the responses tended to flatten out near the "middle-sized" schools. However the Pākehā and "Other" curves were non-decreasing whereas the Māori and Pacific curves had maxima near the middle of the school sizes. In contrast the Asian curve had a *minimum* near the middle. The practical implications of these phenomena are, to say the least, mysterious. A possible (and very tentative) explanation for the local maxima which appear in the curves for Māori and Pacific students might be that as schools grow larger the number of course offerings they provide grows. This permits streaming or tracking, which could create a tendency for Māori and Pacific students to be streamed into less "academic" versions of courses, which are less likely to prepare them properly for attaining NZQA UE. (See Oakes, 1990 for some related ideas.) It is also possible that some sort of reverse of this effect could account for the local minimum in the curve for Asian students. At this time it is not feasible to investigate these conjectures further, nor does it appear profitable to further explore the elaborate model, involving both school size and number of achievement standards available, referred to above.

At this point it is worth noting that one could include all the predictors: school size, decile, number of achievement standards available, and student ethnicity in a single model. Such a model would become unwieldy in its complexity and effectively impossible to interpret. However, it should be noted that if one tests such a model against one which omits the number of achievement standards predictor, one rejects the simpler model with a *p*-value of 5.23×10^{-6}. In other words, there is statistically significant evidence that the number of achievement standards available has predictive power over and above that provided by the other three predictors.

6. DISCUSSION AND CONCLUSIONS

The story that the data are telling us is intricate and complex, but very compelling. The basic question, “Does the number of standards available have an impact upon the number of standards attempted?” is answered in the affirmative, but with the qualifier “but the impact is small”. The amount of impact is so slight as to be an unlikely explanation for the differences between ethnic groups in the number of standards attempted. The question “Does the availability of standards have an effect on the GPA score of students?” is likewise answered in the affirmative, but with the same qualifier.

On the other hand, the question “Does the availability of standards have an effect on the success rate of students in achieving NZQA UE?” is answered in the affirmative, and strikingly so. That is, the more standards that are available, the better students do. Furthermore, the overall impact would appear to be sufficiently large to be of substantial practical interest. The change in the UE success rate, when the number of achievement standards available increases from the mean number (109.4) to the maximum number (166), is estimated to be 10 percentage points, and is at least 7 percentage points (with 95% confidence).

Things become even more interesting when we examine the results in terms of the different ethnic groups (see Figure 8 on page 48). Initially it appears that this positive effect is driven largely by the Pākehā group, with a marginal contribution from the Māori group. Increasing the number of available standards seems to have no effect at all on the Pacific group. This would be disappointing if that were all there was to it: superficially at least it would appear that if steps were taken to make more (achievement) standards available in the lower decile schools, the gap in performance between the Pākehā group and the Māori and Pacific groups would simply be exacerbated. However the story is more intricate than that, and much more hopeful.

The more detailed analysis depicted in Figure 9 on page 49 appears to indicate that increasing the number of standards has a positive impact on the Pacific students of higher academic potential as measured by NCEA GPA, and a *negative* impact on Pacific students with lower ability levels. This is also true of the Māori group.

However, as was pointed out in section 5.6, the negative impacts are not statistically significant. In fact the only ethnicity-GPA category combinations for which these coefficients differ from 0, in a statistically significant manner, are "(Māori, midhi)" and "(Pacific, midhi)". It may seem surprising that when "Pākehā" is subdivided by GPA category none of the individual slope coefficients are significantly different from 0. This phenomenon occurs notwithstanding the fact that when the ethnic groups are *not* subdivided the "slope" coefficient for the "Pākehā" group is strongly significant. What is apparently going on is that for the Pākehā group the impact of availability of standards is relatively small within GPA categories and relatively large between categories.

The important point is that the coefficients are significant and positive *specifically* for Māori and Pacific students in the medium to high ability range. This is exactly the group of students for whom we need and expect to be able to bring about an improvement in academic outcomes. The results of this analysis indicate that such an improvement is possible if more achievement standards are made available to these groups of students.

As a note of caution we should emphasise here that our findings are somewhat in contrast with those of Finn (1999) who found that patterns of course taking were *not* affected by the availability of advanced courses. However it may be the case that our result is not directly comparable with Finn's. Finn is talking about advanced courses; we are considering achievement standards which, while generally considered to be more academically demanding, are not *necessarily* more advanced. Moreover "standard" constitutes a more fine-grained concept than "course". It may be the case that this difference in granularity accounts for at least some of the difference between our findings and Finn's.

Another obvious note of caution relates to the measure of "ability level" that was used in the analysis discussed above. This measure, based on the NCEA GPA, is "post hoc", making the argument we use somewhat circular. It would be much more convincing to have an *a priori* measure of students' ability (based on, say, asTTle, or PAT or MidYIS tests) as the foundation on which to conduct the analysis. Such measures were not available to us and would have required considerable effort to obtain. Gathering such data and incorporating them into analyses similar to those undertaken for this report could constitute an interesting future research project.

The mechanism by which the number of standards available affects the success rate is less than straightforward. As we remarked above, while the number of standards attempted by students does indeed increase (overall) with the number of standards available, the size of the effect compared with the amount of variability about the trend line is minimal. See Figures 2 and 3, pages 42 and 43. It thus seems unlikely that increasing the number of available standards increases success probability simply by virtue of the fact that it tends to increase the number of standards attempted. (In fact a formal hypothesis test of a model involving ethnicity, the number of standards attempted, and the number of standards available, versus a model involving only ethnicity and the number of standards attempted, rejects the reduced model with a p-value of 9.8×10^{-5}.)

Consequently there must be something else going on. We speculate that the number of achievement standards available is to a large extent an indicator of the academic emphasis of a school. In other words, it may be the case that schools which offer more achievement standards are schools that simply place a higher emphasis on, and do a better job of, preparing their students to pursue an academic educational path. Offering more standards may be an indication of how much effort a school makes to meet its students' needs in terms of the goal of achieving NZQA UE. This conjecture is related to the finding of Lee and Bryk (1989) that high levels of achievement are related to the "academic emphasis" of a school.

In respect of the supplementary issue involving the impact of decile on students' success rate, we find that when decile is factored in the relationship between success rate and the number of achievement standards available changes somewhat. The relationship is strongly positive for the Pākehā group and significantly negative for the Pacific group. The relationship is not significant for the other three groups and is "flat" for the Māori group. See Figure 15 on page 55.

The study by Yuan, Turner, and Irving (2010) has revealed (among other findings) that Pacific students in decile 1 schools, on average, do better (in terms of achieving NZQA UE) than Pacific students in decile 4 schools. This appears to confirm the finding of a decreasing relationship between success rate and decile for Pacific students, referred to above.

There is also strong evidence (see Figure 14 on page 54) that the number of achievement standards available increases with decile—by about 10 standards per decile on average. However, despite the link between the school decile and the number of standards available, school decile does not tell the whole story. When models are fitted which include as predictors both decile and the number of standards available (as well as ethnicity) both decile and the number of achievement standards turn out to be significant.

In general, the mechanisms underlying all of the phenomena discussed are unclear. A real understanding of the factors influencing success at achieving NZQA UE would require "teasing out" the detail of the availability (or lack) of achievement standards. It would be interesting to obtain further detail on the *type* of standards that become (increasingly) available as, for instance, decile increases. Questions arise with respect to schools' allocation of effort—are schools where the success rate at achieving NZQA UE is low simply placing emphasis upon goals other than NZQA UE, according to student and parent demand? And is this "demand" real or apparent? This is an important point and could be the subject of a further research project or projects, as could the following questions:

What is the relation between success rate at achieving NZQA UE and success rate at achieving other qualifications (such as level 2 or level 3 certificate)?[9]
How do these issues relate to the fact that the overall Pacific success rate does not increase with increasing availability of achievement standards?

The subjects and standards offered by schools, and the emphasis placed on different sorts of standards, is related to schools' ability to attract appropriately qualified teachers. How might data be obtained to illuminate this issue and connect it to students' success rate in achieving NZQA UE?

In attempting to answer the foregoing questions, the most effective approach might be to focus on individual schools. For instance, this current study may allow Starpath to identify schools which are doing particularly well, i.e. which have a higher than expected success rate (especially for their Māori and Pacific students) given the number of achievement standards that they offer. Such schools could then be studied in more detail with a view to finding out what, if anything, they do that is "special" and to learn from their success.

The issue of the direction of causation ("Does increasing the number of available achievement standards *cause* the success rate to increase, or is there perhaps an effect by which having successful students at a school causes more achievement standards to be made available?") may be deemed to be still open to some debate. However, the argument given in section 5.7, while perhaps not settling the debate, provides fairly convincing evidence that the direction of causation is as we are assuming it to be.

[9] We do not currently have data with which to address this question.

It might be of interest to include a random effect (see for example Pinheiro and Bates, 2000) for “school” in some of the models fitted for this report. Given the complexity of the data and the general complexity of mixed effects modelling, this would be a substantial undertaking. There are also issues to be addressed in respect of the schools in the study being a “random sample”. Examining the implications of incorporating a random effect for “school” could constitute a future research project.

Further technical issues in respect of the statistical analysis include the probable presence of extrabinomial variance in the success rates. More complex models allowing for over-dispersion (fitted via quasi-likelihood) could be explored.

In summary, the main finding of this report is that increasing the number of achievement standards available has an intricate effect on student success rate when ethnicity is taken into consideration. If we refine the analysis of the effect according to the academic “ability range” as well as ethnicity, we find a positive effect in the “midhi” range for both Māori and Pacific students. This is excellent news. It indicates that there is a solid chance of improvement in the success rates of Pacific and Māori students with “reasonably high” academic potential, which is precisely where one would hope for there to be a window of opportunity.

7. REFERENCES

Ayalon, H., & Gamoran, A. (2000). Stratification in academic secondary programs and educational inequality in Israel and the United States. *Comparative Education Review 44*(1), 54–80.

Finn, J. D. (1997). Course taking in high school: Is opportunity equal? spotlight on student success. Technical Report 301, Mid-Atlantic Laboratory for Student Success, Philadelphia, PA.

Finn, J. D. (1999). Opportunity offered – opportunity taken: Course taking in american high schools. *ETS Policy Notes 9*(1), 2–10.

Finn, J. D., Gerber, S. B., & Wang, M. C. (2001). Taking mathematics in high school: Is opportunity equal? Technical Report 600, Mid-Atlantic Laboratory for Student Success, Philadelphia, PA.

Finn, J. D., Gerber, S. B., & Wang, M. C. (2002). Course offerings, course requirements, and course taking in mathematics. *Journal of Curriculum and Supervision 17*(4), 336–366.

Hipkins, R., & Vaughan, K. (2002). *Learning curves: Meeting student needs in an evolving qualifications regime. From cabbages to kings: A first report*. Wellington, New Zealand: NZCER.

Hipkins, R., Vaughan, K., Beals, F., & Ferral, F. (2004). *Learning curves: Meeting student learning needs in an evolving qualifications regime. Shared pathways and multiple tracks: A second report*. Wellington, New Zealand: NZCER.

Hipkins, R., Vaughan, K., Beals, F., Ferral, H., & Gardiner, B. (2005). *Shaping our futures: Meeting secondary students' learning needs in a time of evolving qualifications. Final report of the Learning Curves project*. Wellington, New Zealand: NZCER.

Lee, V. E. (1993). Educational choice: The stratifying effects of selecting schools and courses. *Educational Policy 7*(2), 125–148.

Lee, V. E., & Bryk, A. S. (1988). Curriculum tracking as mediating the social distribution of high school achievement. *Sociology of Education 61*(2), 78–94.

Lee, V. E., & Bryk, A. S. (1989). A multilevel model of social distribution of high school achievement. *Sociology of Education 62*(3), 172–192.

Lee, V. E., Croninger, R. G., & Smith, J. B. (1997). Course-taking, equity, and mathematics learning: Testing the constrained curriculum hypothesis in U.S. secondary schools. *Educational Evaluation and Policy Analysis 19*(2), 99–121.

Madjar, I., McKinley, E., Jensen, S. & vane der Merwe, A. (2009). *Towards university: Navigating NCEA course choices in low-mid decile schools.* Starpath Project, The University of Auckland (www.starpath.auckland.ac.nz)

Meyer, L. H., Weir, K., McClure, J., Walkey, F., & McKenzie, L. (2007). *Final report: Longitudinal research on the relationship between the NCEA and student motivation and achievement*. Wellington, New Zealand: Victoria University of Wellington.

Meyer, L., McClure, J., Walkey, F., McKenzie, L., & Weir, K. (2006). *Final report: The impact of the NCEA on student motivation*. Wellington, New Zealand: Victoria University of Wellington.

Oakes, J. (1990). *Multiplying Inequalities: The Effects of Race, Social Class, and Tracking on Opportunities to Learn Mathematics and Science*. Santa Monica, CA: Rand Corp.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-Effects Models in* S *and* S-PLUS. New York: Springer.

Shulruf, B., Hattie, J., & Tumen, S. (2007). The predictability of enrolment and first year university results from secondary school performance. Technical Report 7, Starpath Project, University of Auckland.

Turner, T. R. (2007). The effect of conditioning upon success on the approved subject criterion. Technical Report 19, Starpath Project, University of Auckland.

Turner, T. R., Li, M., & Yuan, J. (2010). The impact of a revised UE criterion upon certain subgroups of students. Technical report, Starpath Project, University of Auckland. (In progress.).

Yuan, J., Turner, T. R., & Irving, S. E. (2010). Factors influencing University Entrance success rate. Technical report, Starpath Project, University of Auckland. (In progress.).

APPENDIX 1: Looking for Influential Points

It was suggested to us that the unexpected result in respect of Pacific students (i.e. that number of standards available had no influence on their performance) might be driven by a small number of possibly large schools. To investigate this possibility we selected out the data pertaining to Pacific students only, and then eliminated schools where student numbers were "small", i.e. those schools in which there were fewer than 20 (Pacific) students in Year 13 or higher. Twenty schools remained. For each of these we calculated raw success probability estimates as the number achieving NZQA UE divided by the total number. We then fitted a binomial logistic model (predicting NZQA UE from number of achievement standards available) and calculated the model-based success probability estimates as the predicted values ("type=response") from the fit. Finally we calculated pseudo-residuals equal to the raw success probability estimates minus the model based estimates.

These residuals were plotted against the predictor, as shown in Figure 18 on page 58. Since different points have different amounts of uncertainty associated with them (the smaller the number of students, the greater the uncertainty) we plotted the points as discs having radii proportional to the reciprocal of the associated "standard error". (Thus the smaller the plotting symbol, the greater the uncertainty.) The standard error was calculated as $\sqrt{\hat{p}_i(1-\hat{p}_i)/n_i}$ where p_i is the fitted value from the model and n_i is the number of students in the sample, for the i_{th} school.

The horizontal lines in Figure 18 are at ± 2 *SE* where *SE* is the mean of the standard errors discussed above. There are two (substantially "large") points above the upper line, which might thus be deemed outliers. When the model was refitted with these two outliers removed from the data set, the *p*-value of the predictor decreased from 0.96 to 0.27. However the coefficient for the predictor was "even more negative" in the second instance, so eliminating the possible outliers makes the mystery *deeper* than it was, rather than less so.

APPENDIX 2: Looking for a Threshold

It was suggested to us that, in view of the strong evidence of an overall impact of quantity of standards on success rate, there might be some sort of “threshold” or “tipping point” (in the number of standards available) above which the success rate improves dramatically. To investigate this question we calculated the success rate amongst all students at schools where the number of achievement standards available was at least “Θ” where Θ represents a generic cut-off value. We also calculated the success rate amongst all students at schools where the number of achievement standards available was at *most* Θ. We did this for various values of Θ and plotted the two success rates against Θ. The result is shown in Figure 19 on page 59. In the first plot there is a fairly prominent threshold around $\Theta = 120$. In the second there is a similarly prominent threshold around $\Theta = 45$.

This would appear to be saying that if a school wants to do “well” it should offer at least 120 Level 3 achievement standards, and if it wants to avoid doing “really badly” it must offer at least 45 Level 3 achievement standards. It is a bit disappointing, though perhaps not surprising, that the upper threshold is so high; it will (or would) take substantial effort to raise the number of available standards to that level at the schools which fall below that level. Also it must be remembered that the effect of the number of achievement standards on success rate is largely driven by the Pākehā ethnic group — which is not the group causing particular concern. Nevertheless the existence of these thresholds is interesting and may have important practical implications.

APPENDIX 3: Checking for a Sex Effect

It was suggested to us that a sex effect might have an impact upon the results, and so we investigated this possibility. As is well known, the sex factor does indeed have an impact on academic achievement, with females consistently outperforming males. Our analysis indicates that Māori boys are not outperformed by Māori girls to any greater extent than Pākehā boys are outperformed by Pākehā girls, and similarly for Asians and "Others". In contrast, the extent by which Pacific boys are outperformed by Pacific girls is significantly greater than is the case for the Pākehā group (p-value = 0.007).

Our real concern, however, is with the response to the "continuous" predictor, i.e. the number of achievement standards available, and we find that this response is apparently unaffected by sex. That is, testing for a sex by ethnicity by continuous predictor interaction fails to reject the null hypothesis (of no such interaction) with a p-value of 0.5. Visual confirmation of this conclusion is provided by Figure 20 on page 60, in which we see that the patterns of the curves differ strongly between ethnic groups, but not between sexes within ethnic groups.

APPENDIX 4: Comparison of Responder and Non-responder Schools

We compared the two categories of schools with respect to:

- size of the school, i.e. number of students in year 13 or higher who attended each school and had Level 3 NCEA results
- ethnic mix
- number of achievement standards attempted by students
- NCEA GPA attained by students
- fraction of students (in Year 13 or higher) achieving NZQA UE, sub-classified by ethnicity.

The results are shown in Figures 21 to 29 on pages 61 to 69. Although there are clearly “differences” between the two categories of schools, nothing (to our minds) leaps out as indicating differences that could bias our conclusions in any appreciable way.

APPENDIX 5: TABLES

Table 1: Sample means of the quantity of standards attempted, followed by 95% confidence intervals for the true means

	All standards		**Approved standards**	
Ethnicity	By count	By credits	By count	By credits
Pākehā	23.5; [23.4, 23.7]	95.2; [94.7, 95.6]	20.7; [20.6, 20.9]	83.4; [82.9, 83.9]
Māori	18.9; [18.6, 19.2]	75.8; [74.6, 77.0]	14.7; [14.4, 15.0]	59.3; [58.1, 60.5]
Pacific	18.7; [18.4, 19.0]	74.8; [73.6, 76.1]	13.8; [13.5, 14.1]	55.6; [54.3, 57.0]
Asian	25.0; [24.8, 25.3]	102.3; [101.4, 103.2]	22.9; [22.7, 23.2]	91.7; [90.7, 92.7]
Other	22.0; [21.3, 22.7]	87.3; [84.6, 90.0]	18.9; [18.1, 19.6]	75.1; [72.2, 78.1]

Table 2: Expected change in success probability, with confidence intervals and percentage change (entire data set)

	Pākehā	**Māori**	**Pacific**	**Asian**	**Other**
Change	0.1178	0.0612	-0.0010	0.0314	0.0766
95% CI:	[0.08,0.15]	[0.00+,0.12]	[-0.04,0.04]	[-0.01,0.08]	[-0.04,0.20]
Percent change:	23.8	21.1	-0.5	6.0	26.7

Note: The “0.00+” in this table indicates a value which is strictly positive but is zero when rounded to two decimal places.

Table 3: Expected change in success probability, with confidence intervals and percentage change (restricted data set)

	Pākehā	**Māori**	**Pacific**	**Asian**	**Other**
Change	0.0764	0.0619	0.0747	0.0227	0.0275
95% CI:	[0.04,0.11]	[-0.01,0.14]	[-0.01,0.15]	[-0.02,0.07]	[-0.10,0.15]
Percent change:	11.0	11.4	17.2	3.3	5.5

APPENDIX 6: FIGURES

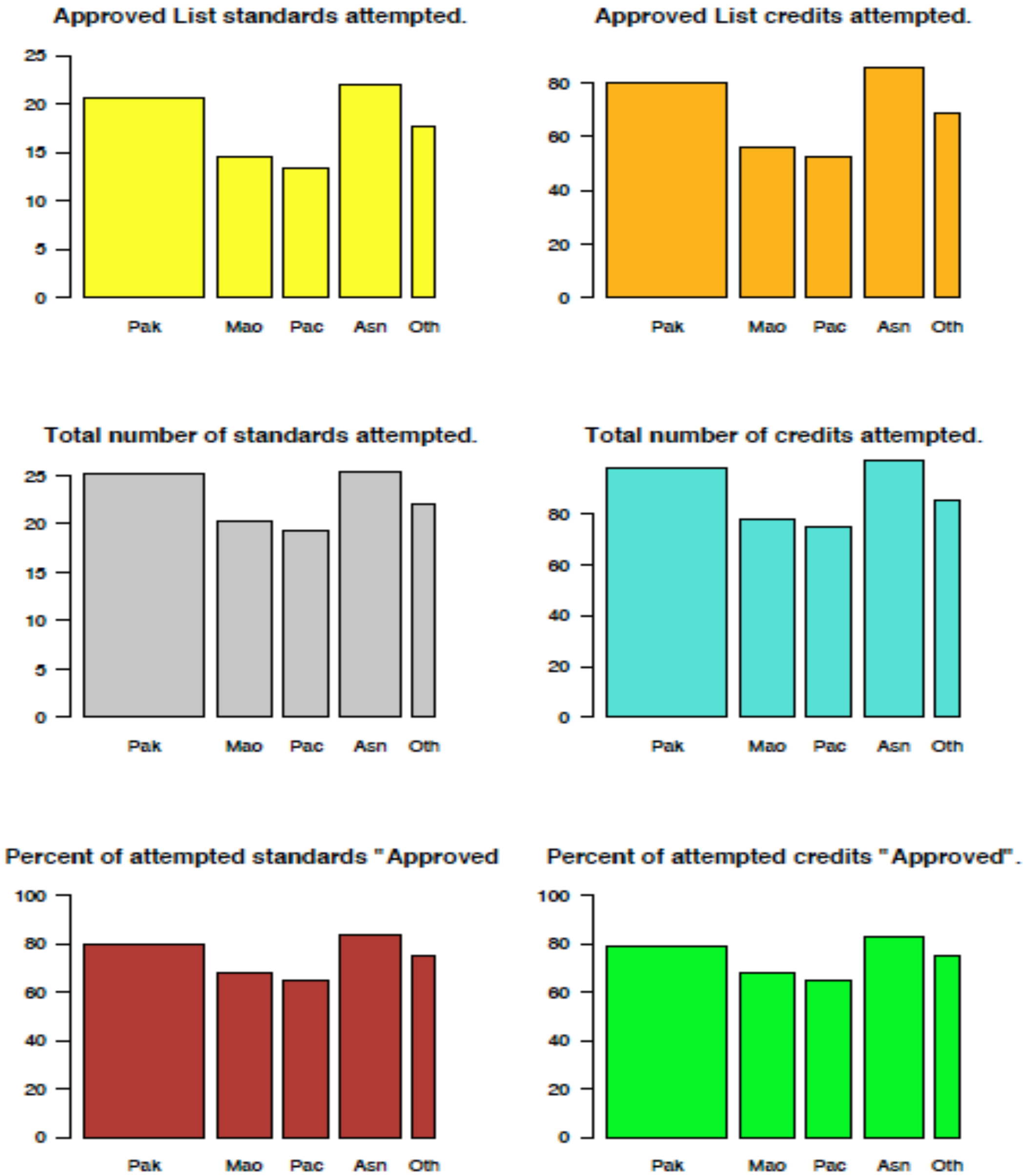

Figure 1: Bar plots of the "quantity" of standards attempted

Note: The widths of the bars in these bar plots are proportional to the square roots of the corresponding sample sizes. This convention is considered to be the best way of conveying to the viewer the relative sizes of the samples being compared.

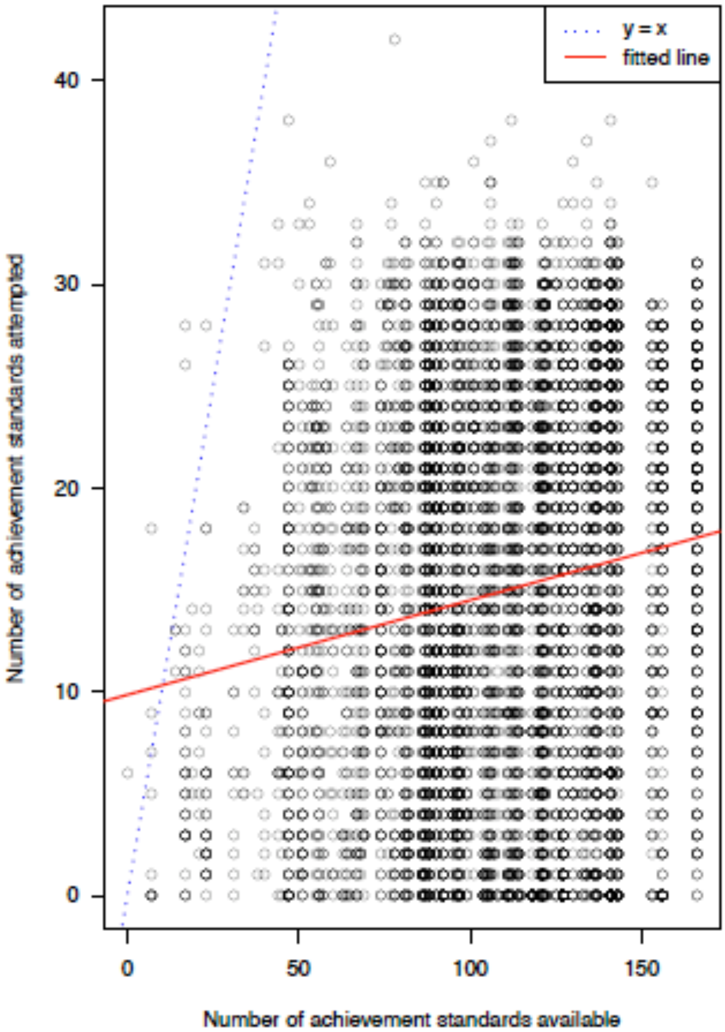

Figure 2: Number of achievement standards attempted versus number available

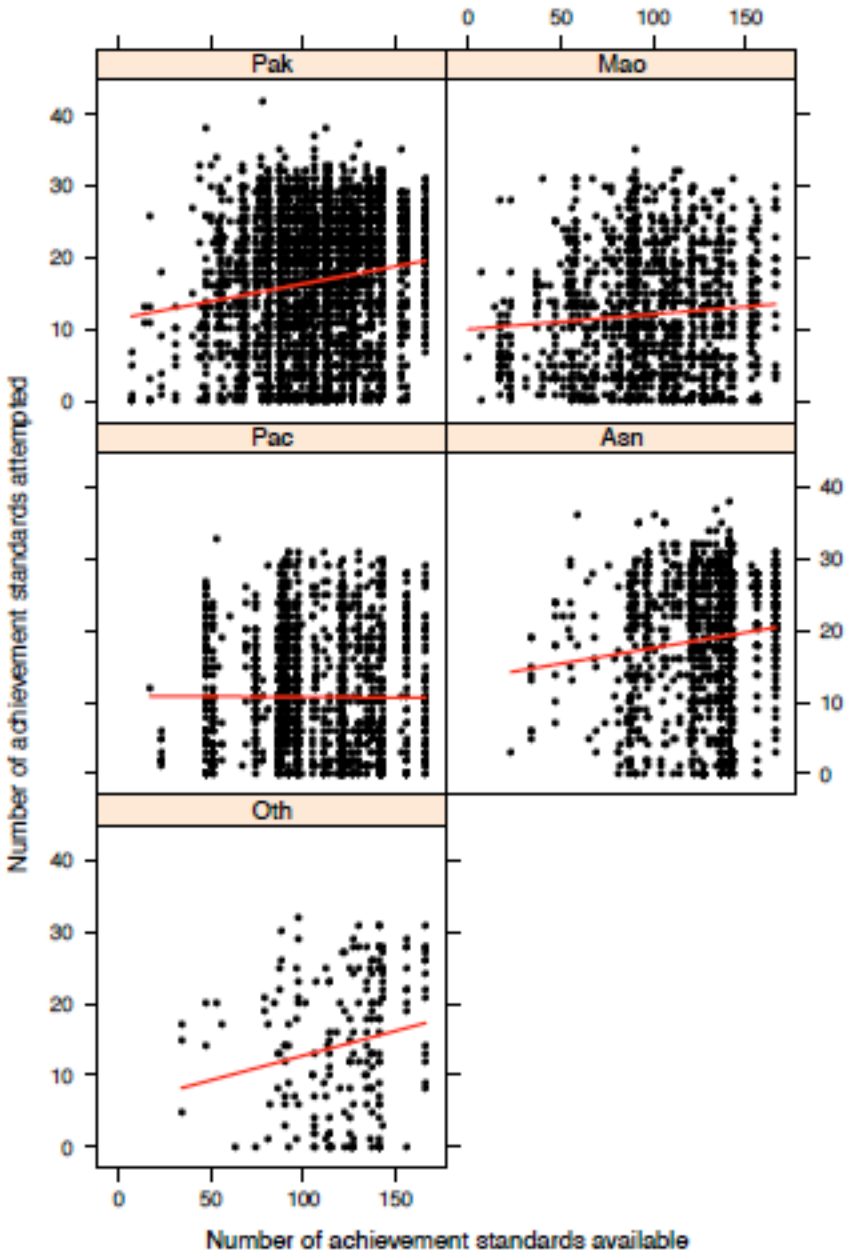

Figure 3: Number of achievement standards attempted versus number available, by ethnicity

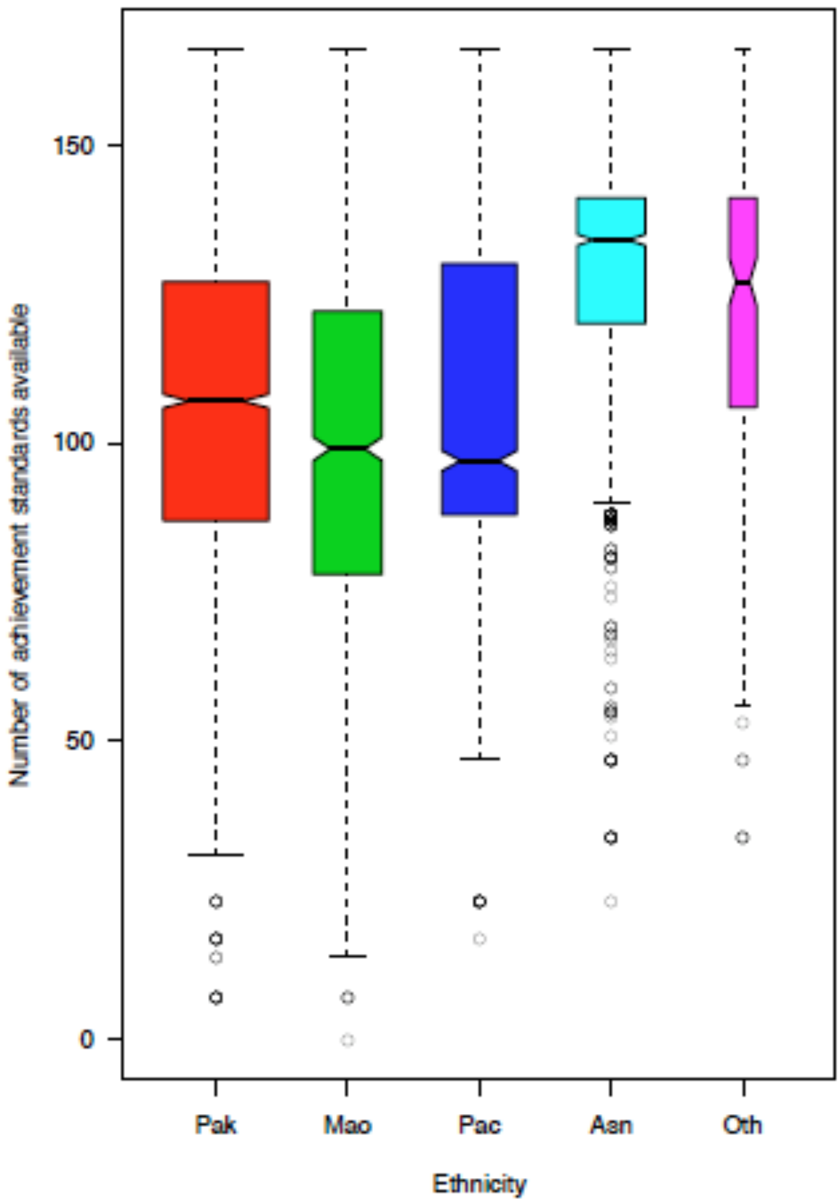

Figure 4: Box-and-whisker plots of the numer of achievement standards available, by ethnicity

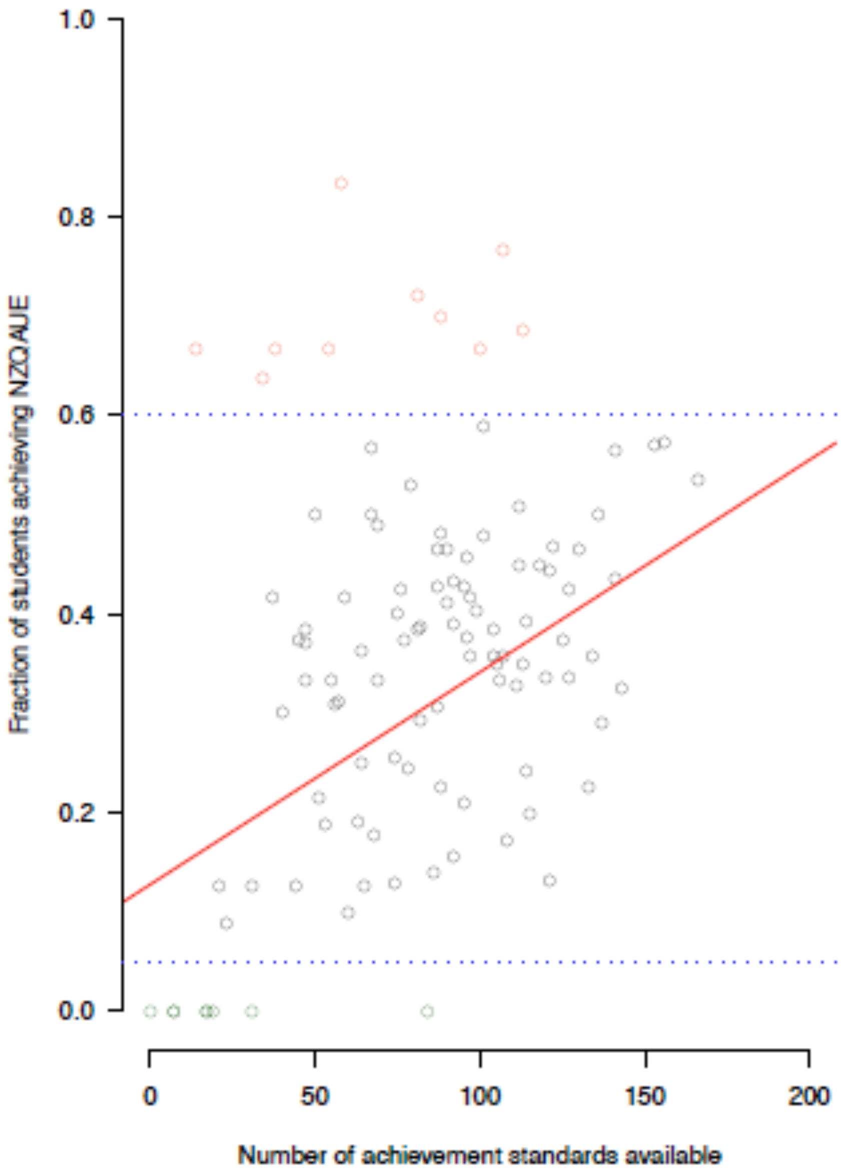

Figure 5: Success fraction versus number of achievement standards available

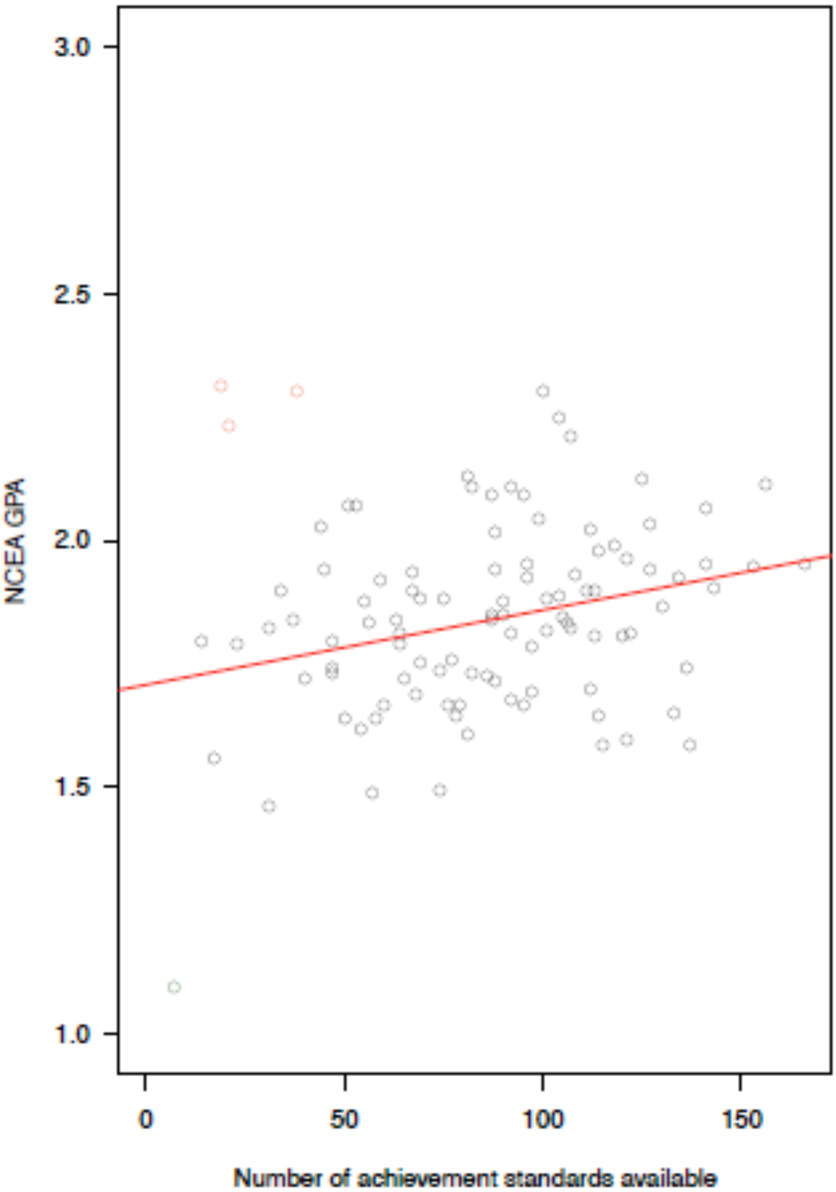

Figure 6: NCEA GPA versus number of achievement standards available, school-by-school analysis

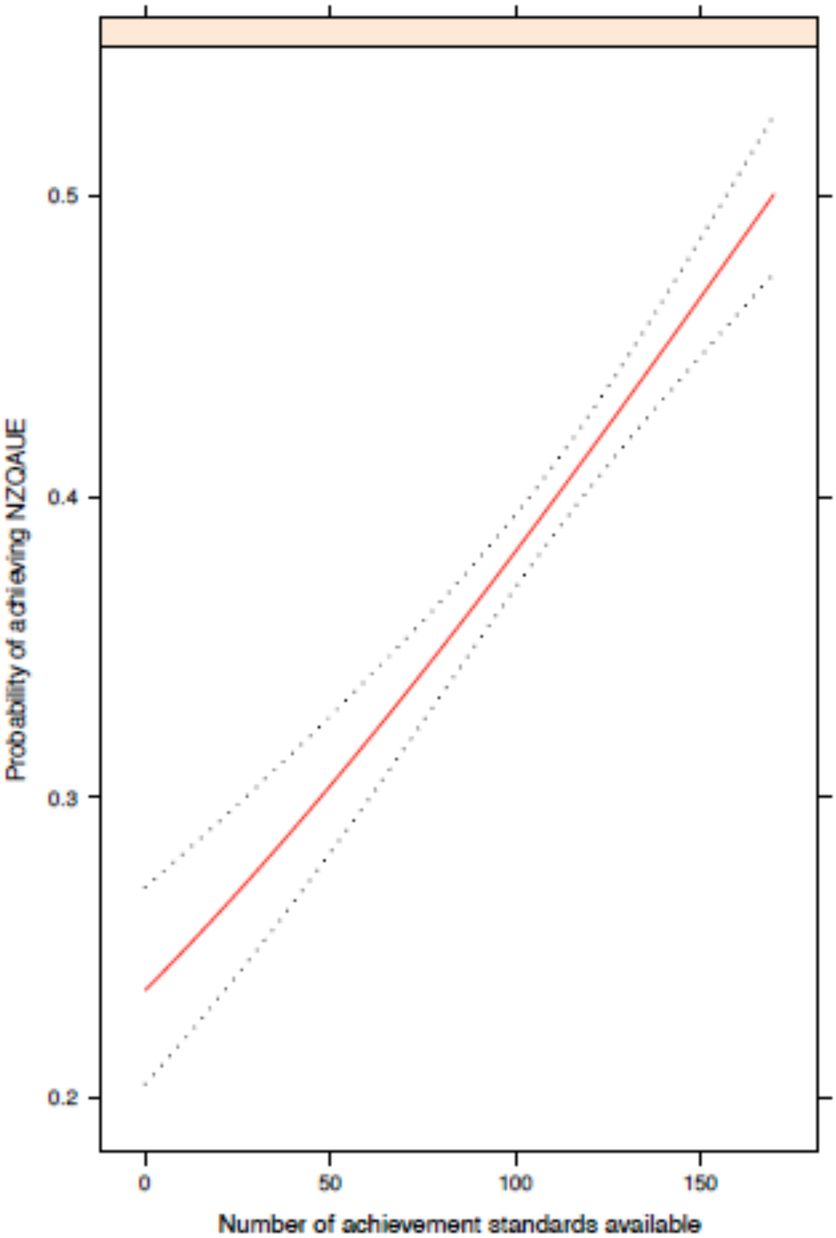

Figure 7: Probability of achieving NZQA UE versus number of achievement standards available

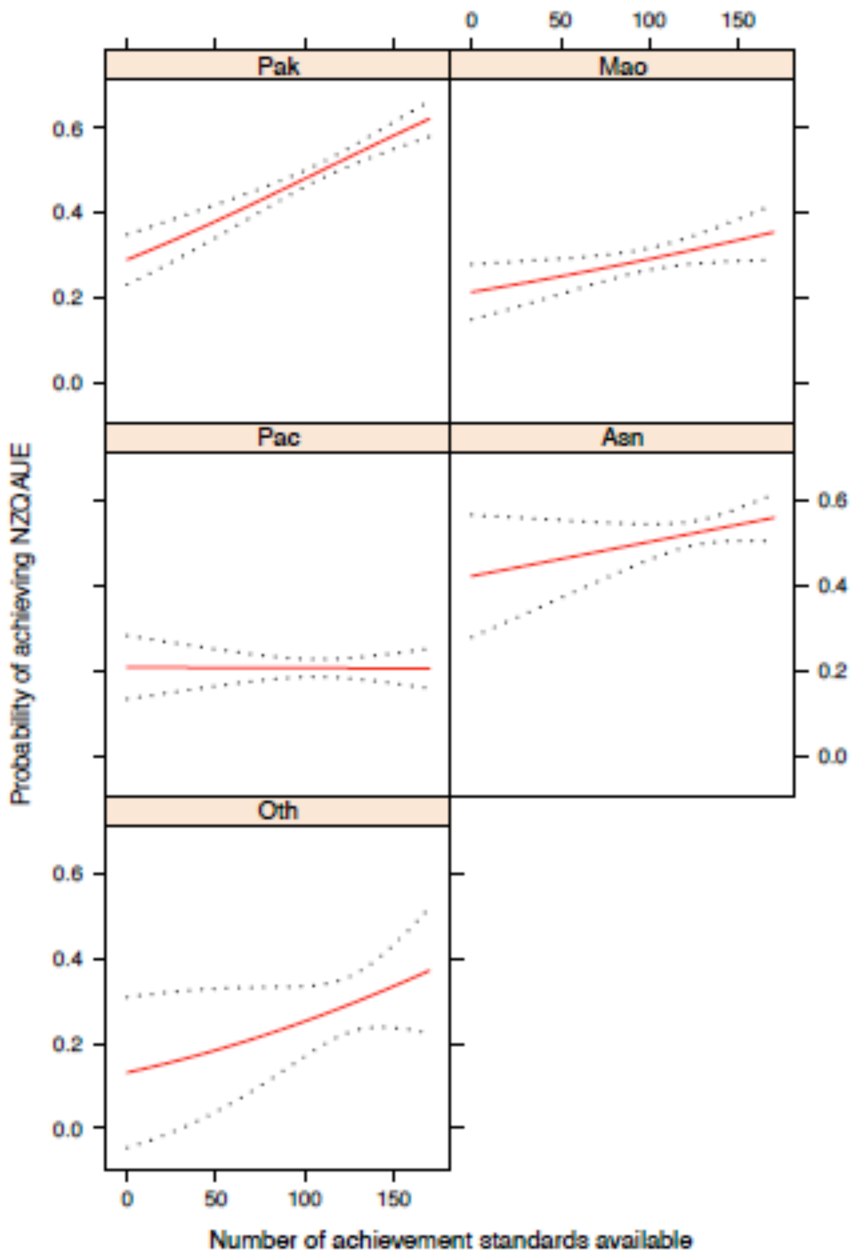

Figure 8: Probability of achieving NZQA UE versus number of achievement standards available, by ethnicity

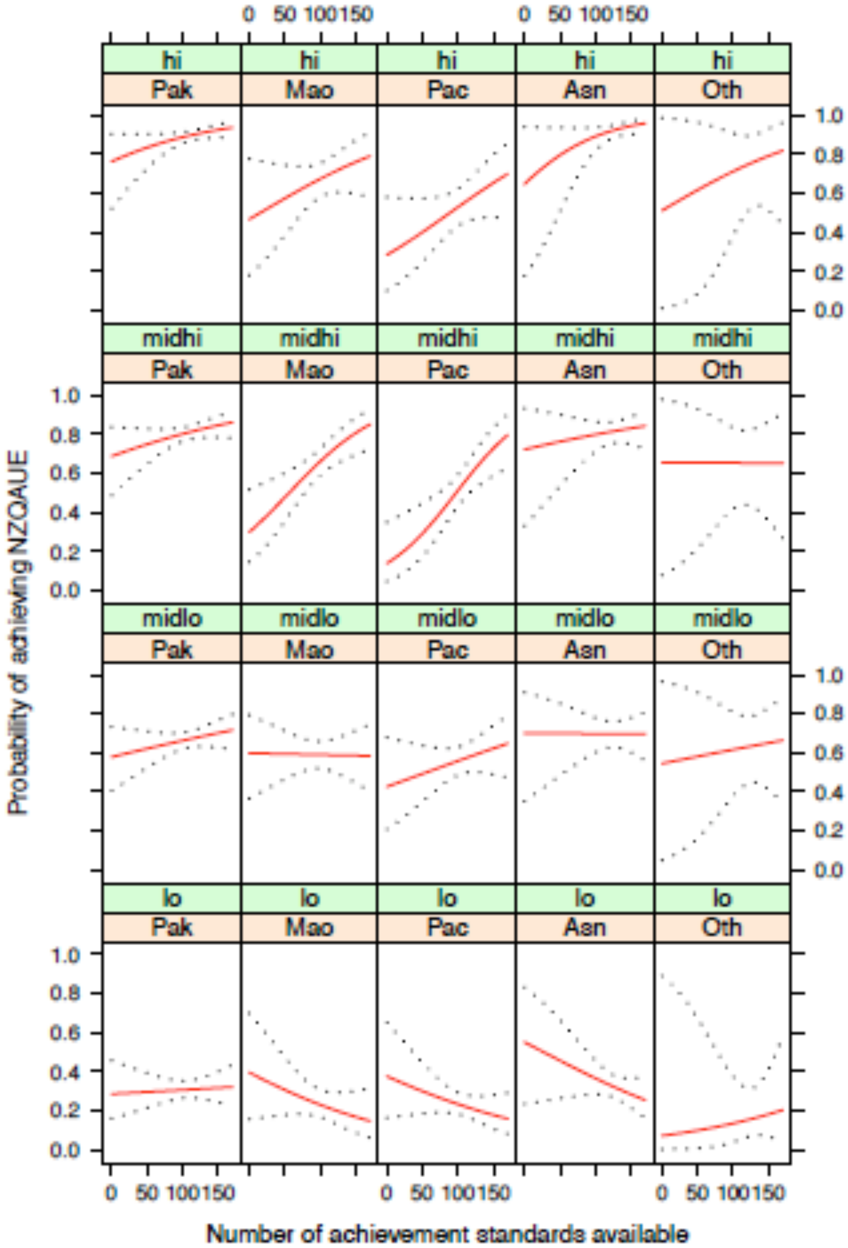

Figure 9: Probability of achieving NZQA UE versus number of achievement standards available, by ethnicity and NCEA GPA category

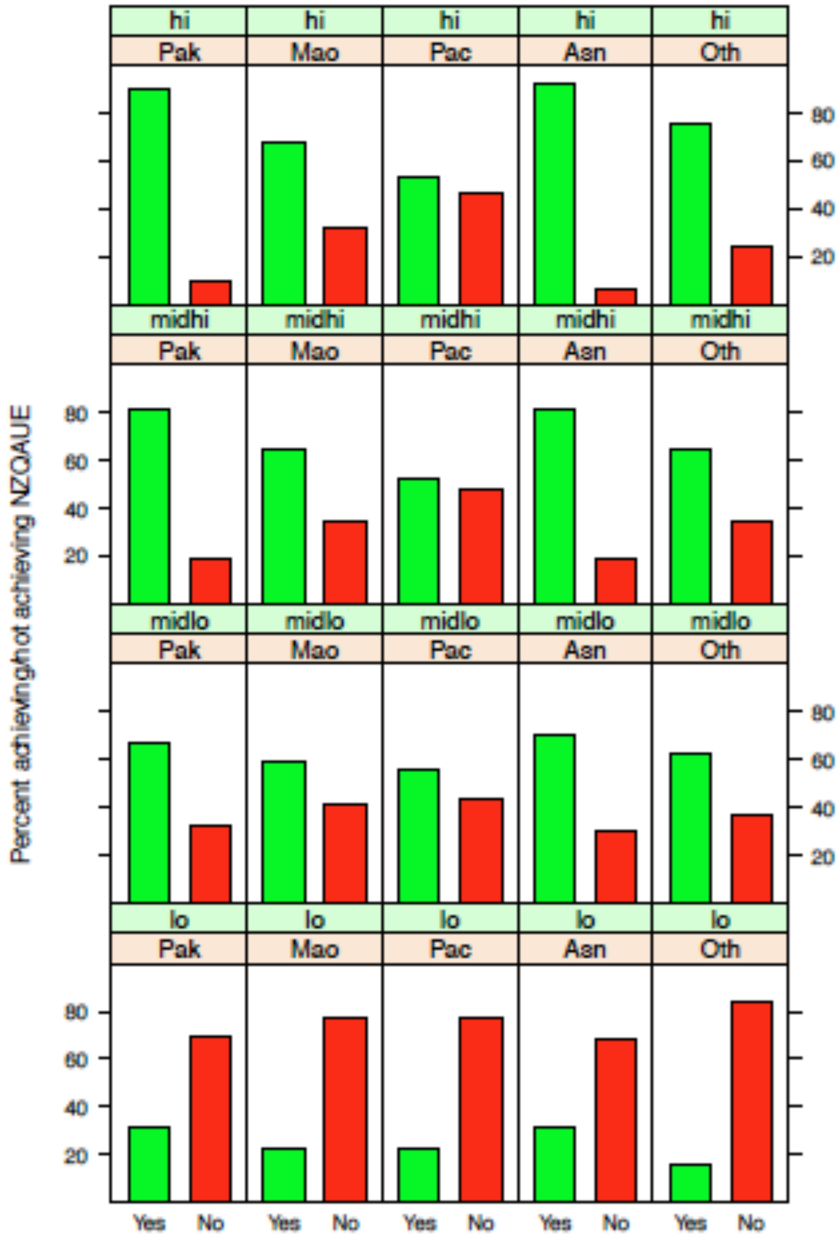

Figure 10: Percentage achieving NZQA UE, by ethnicity and NCEA GPA category

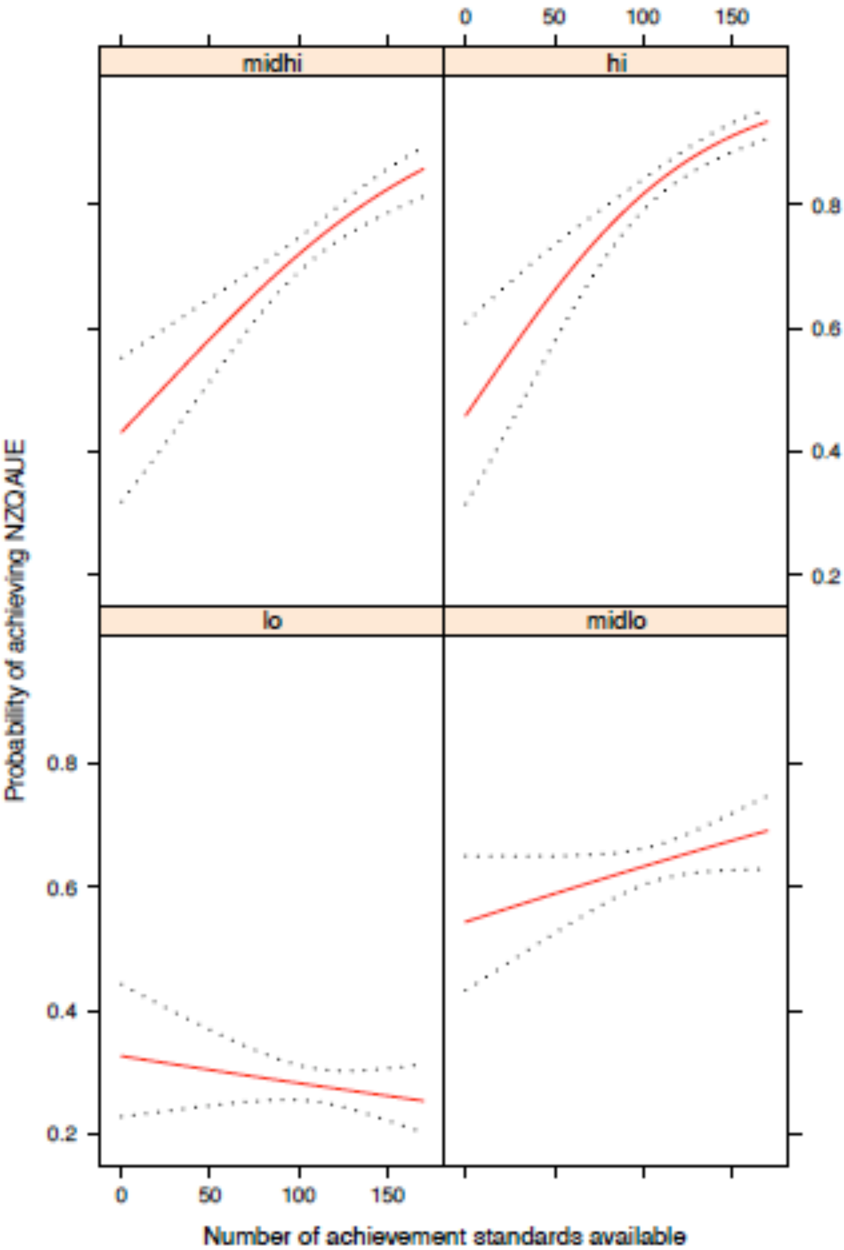

Figure 11: Probability of achieving NZQA UE versus number of achievement standards available, by NCEA GPA category

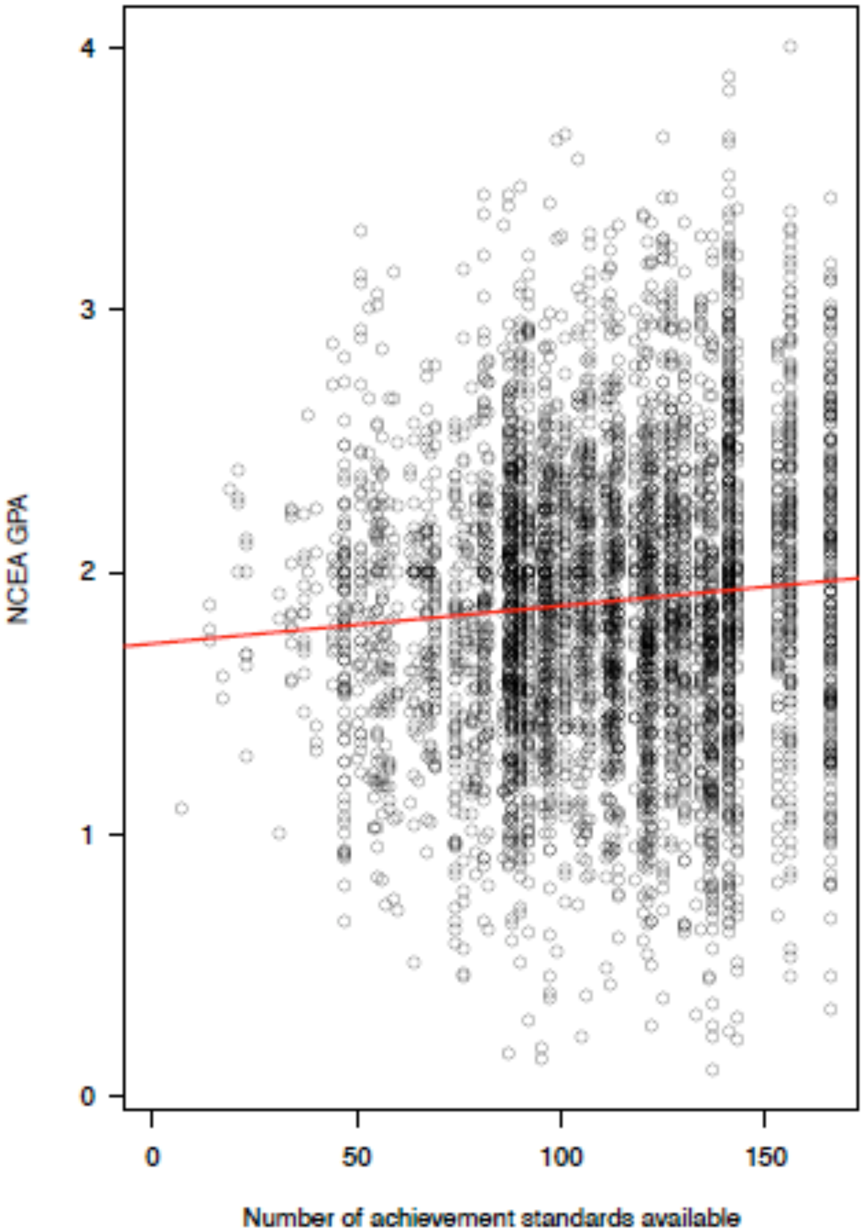

Figure 12: NCEA GPA versus number of achievement standards available, student-by-student analysis

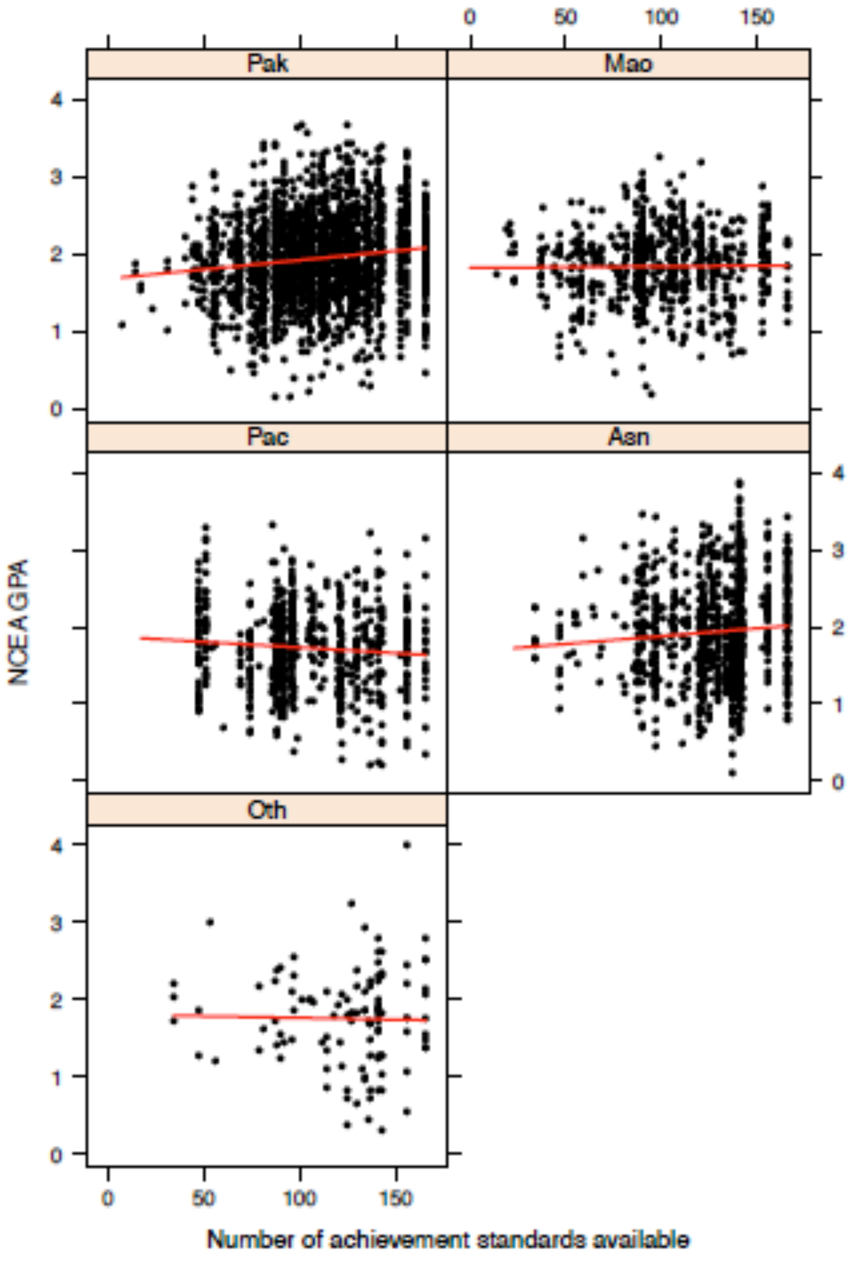

Figure 13: NCEA GPA versus number of achievement standards available, by ethnicity

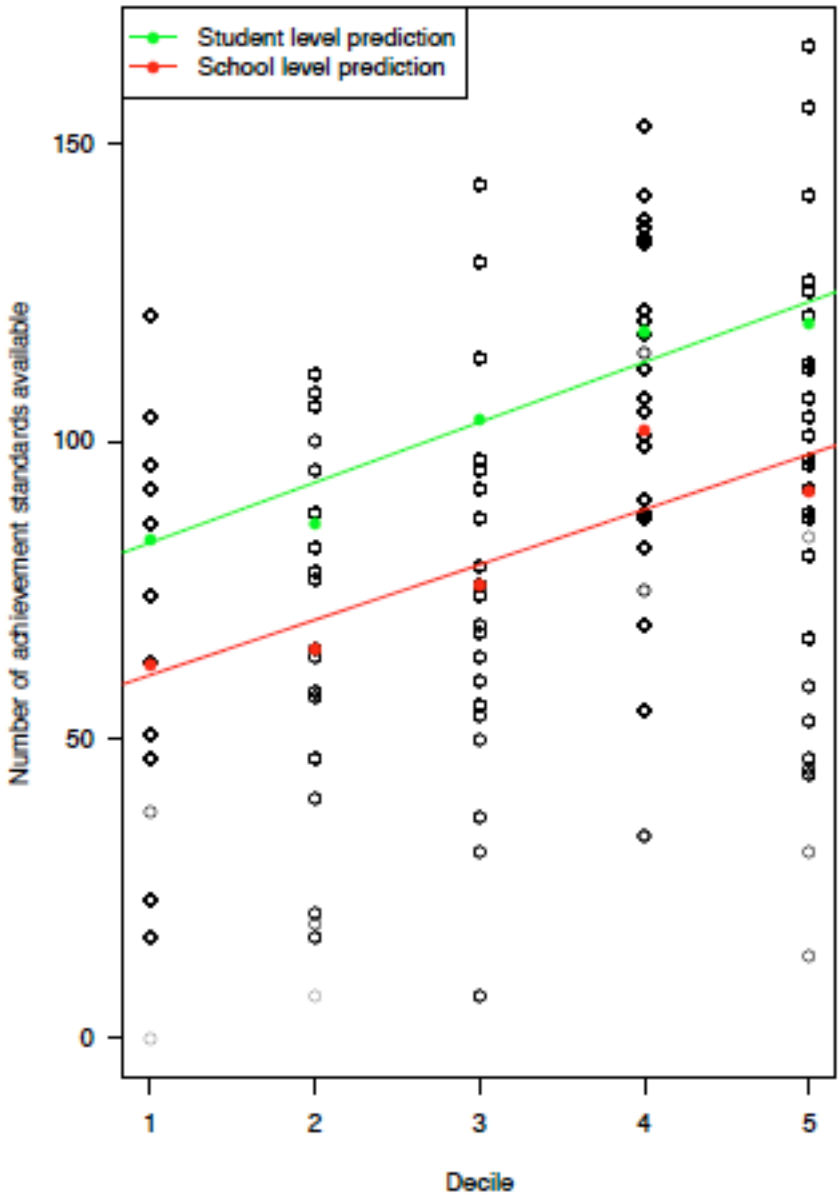

Figure 14: Number of achievement standards available, versus decile

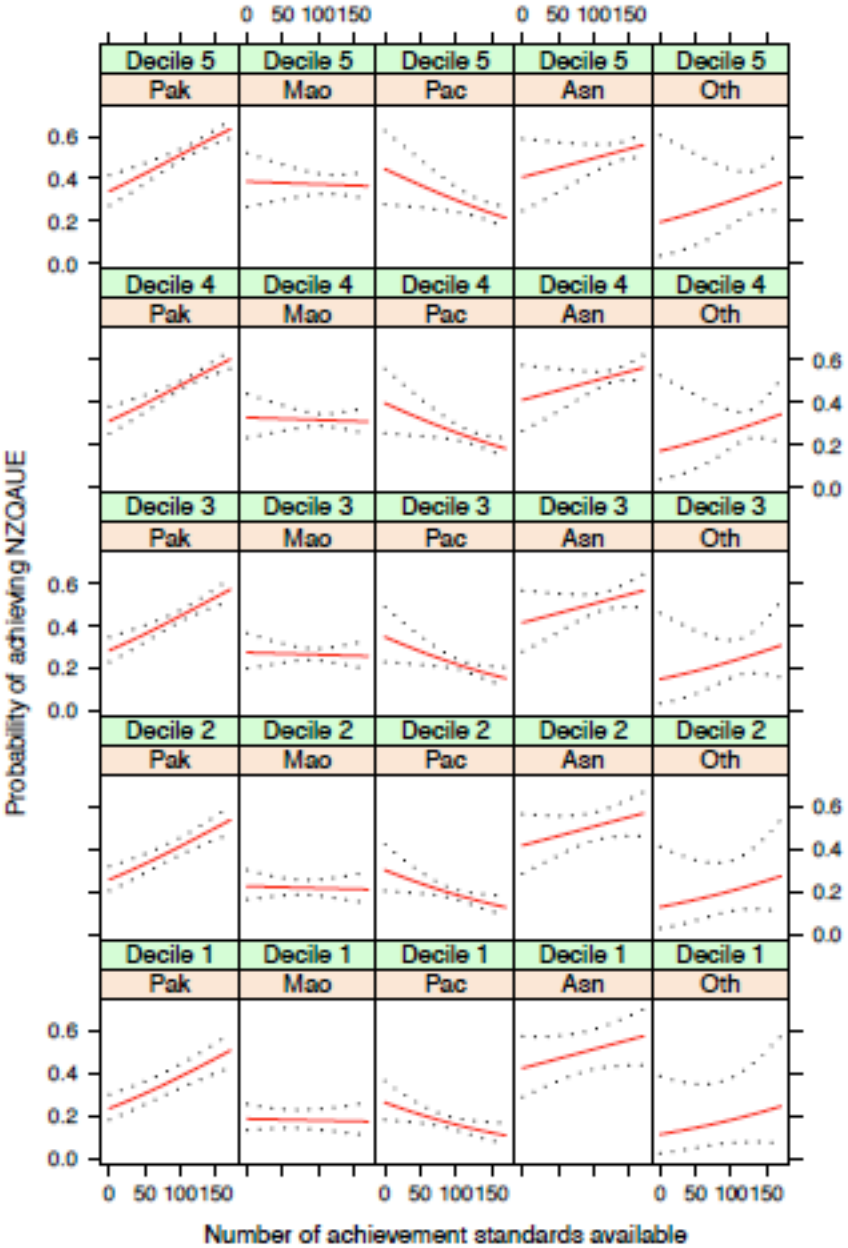

Figure 15: Probability of achieving NZQA UE versus number of achievement standards available, by ethnicity and decile

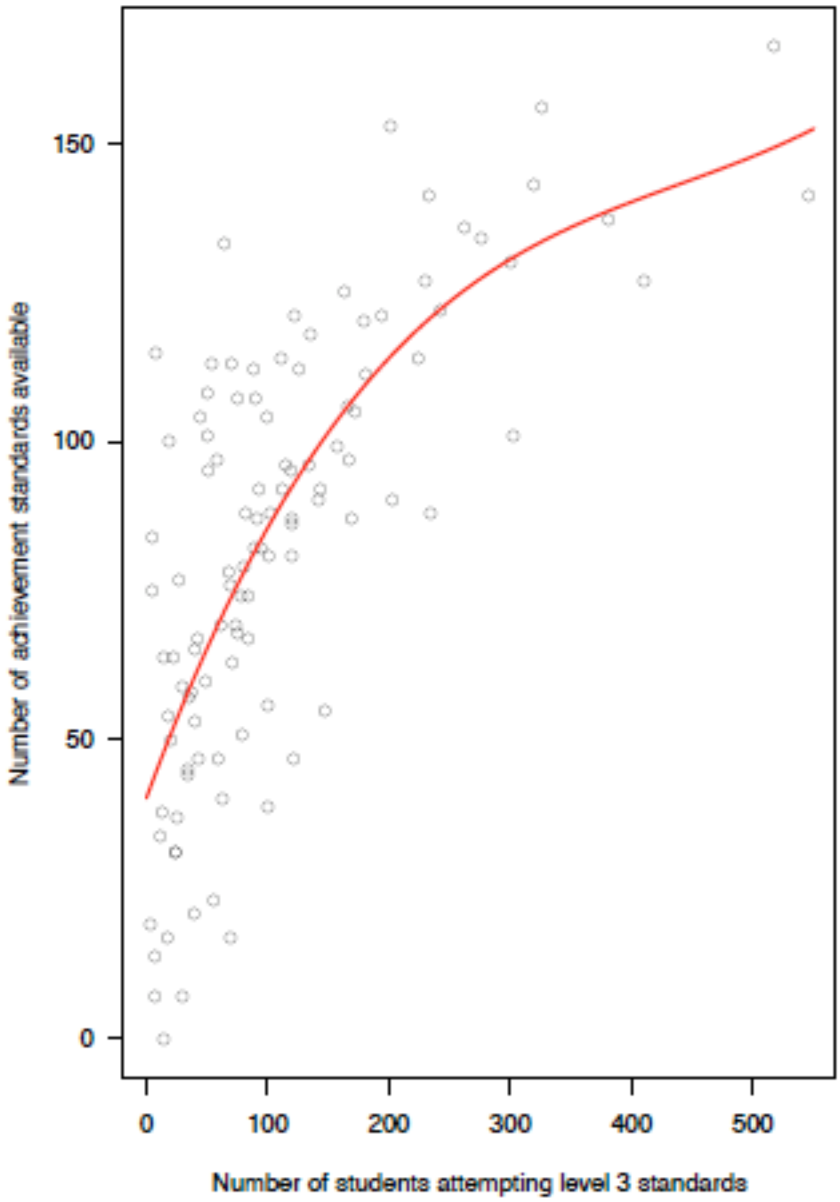

Figure 16: Number of achievement standards available versus a measure of school size

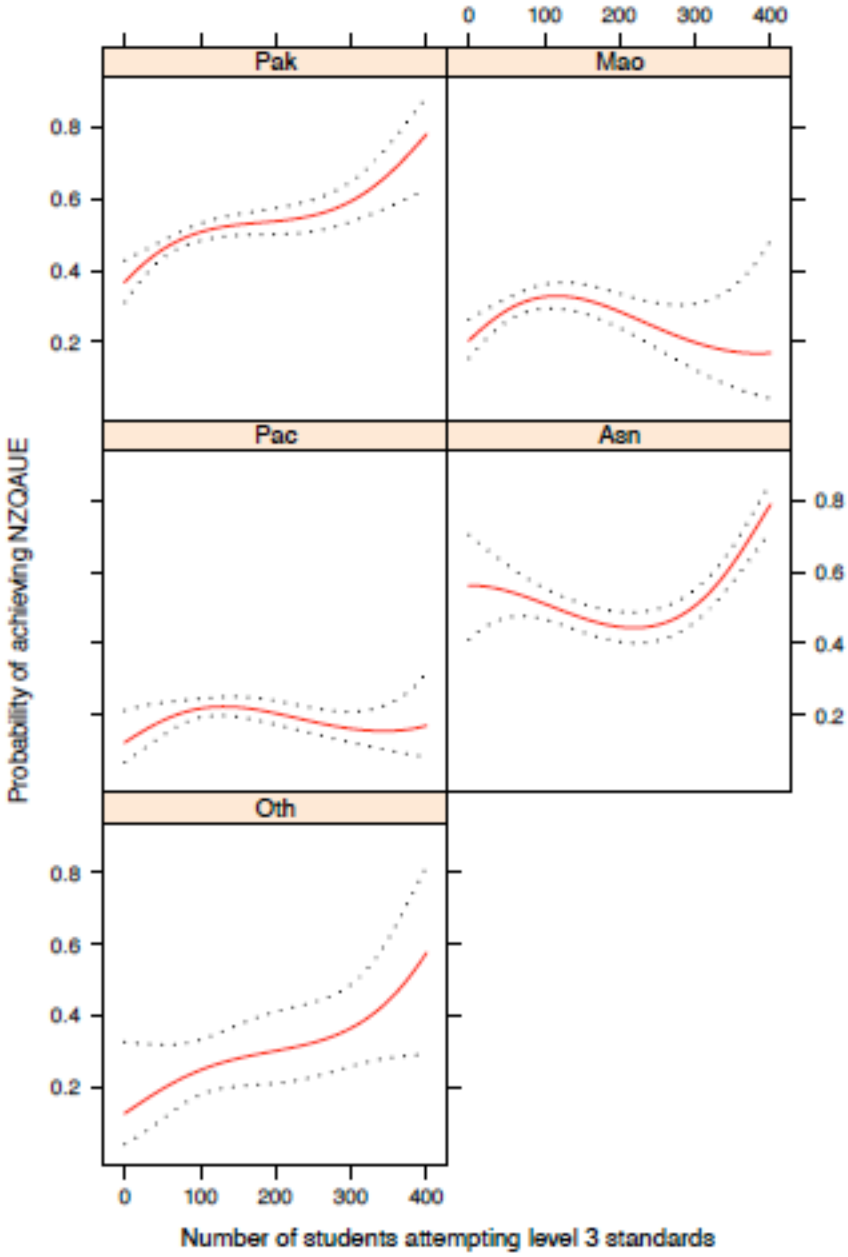

Figure 17: Probability of achieving NZQA UE versus a measure of school size, by ethnicity

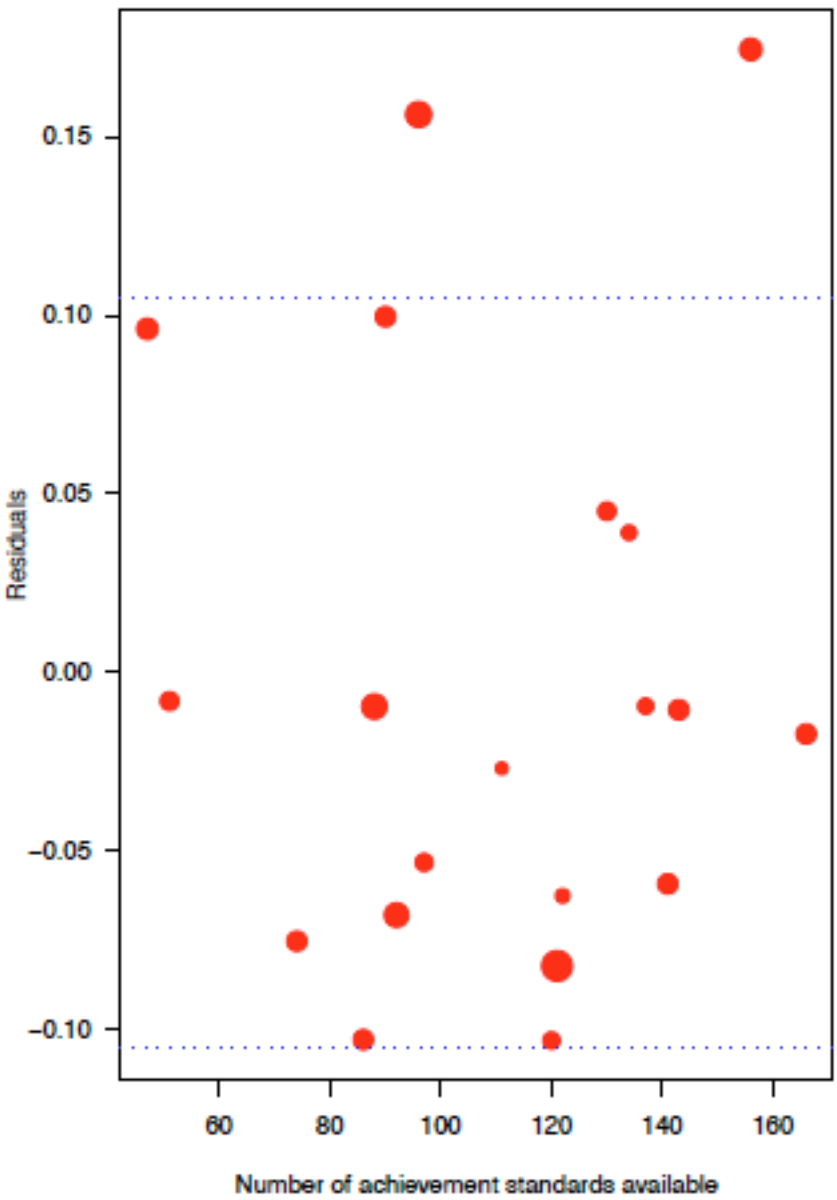

Figure 18: A residual plot which seeks to identify influential schools

Note: The sizes of the dots are proportional to their “reliability”.

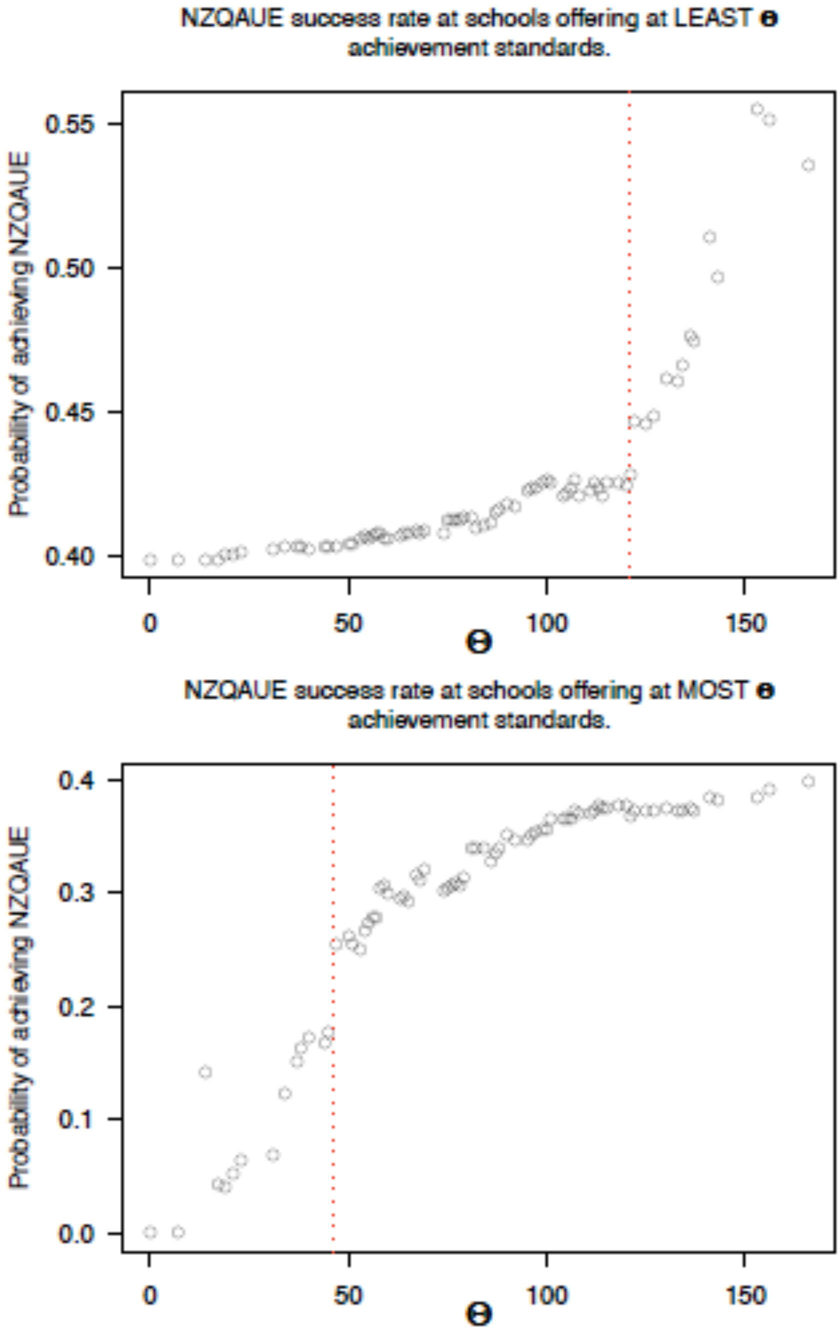

Figure 19: Success thresholds in terms of the number of achievement standards offered

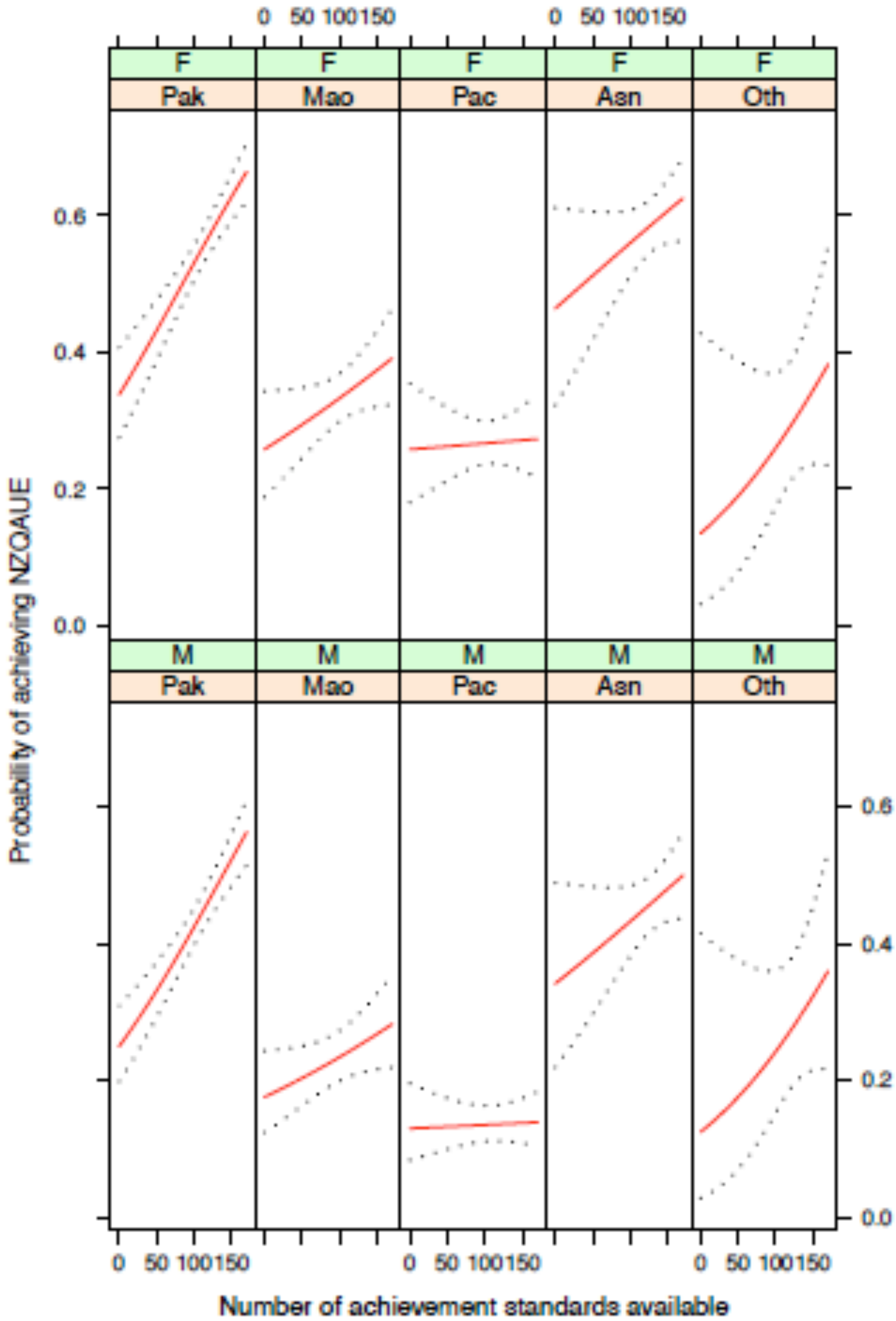

Figure 20: Probability of achieving NZQA UE versus number of achievement standards available, by sex and ethnicity

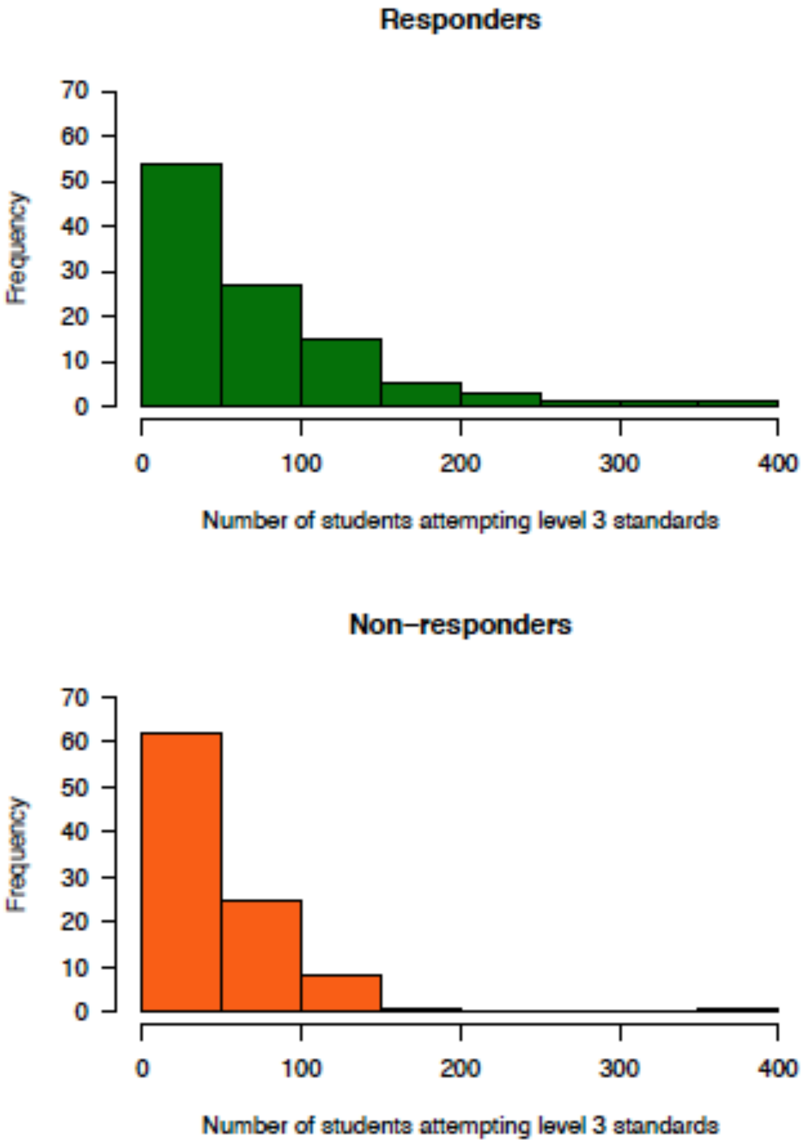

Figure 21: Histograms of a measure of school size, comparing responders and non-responders

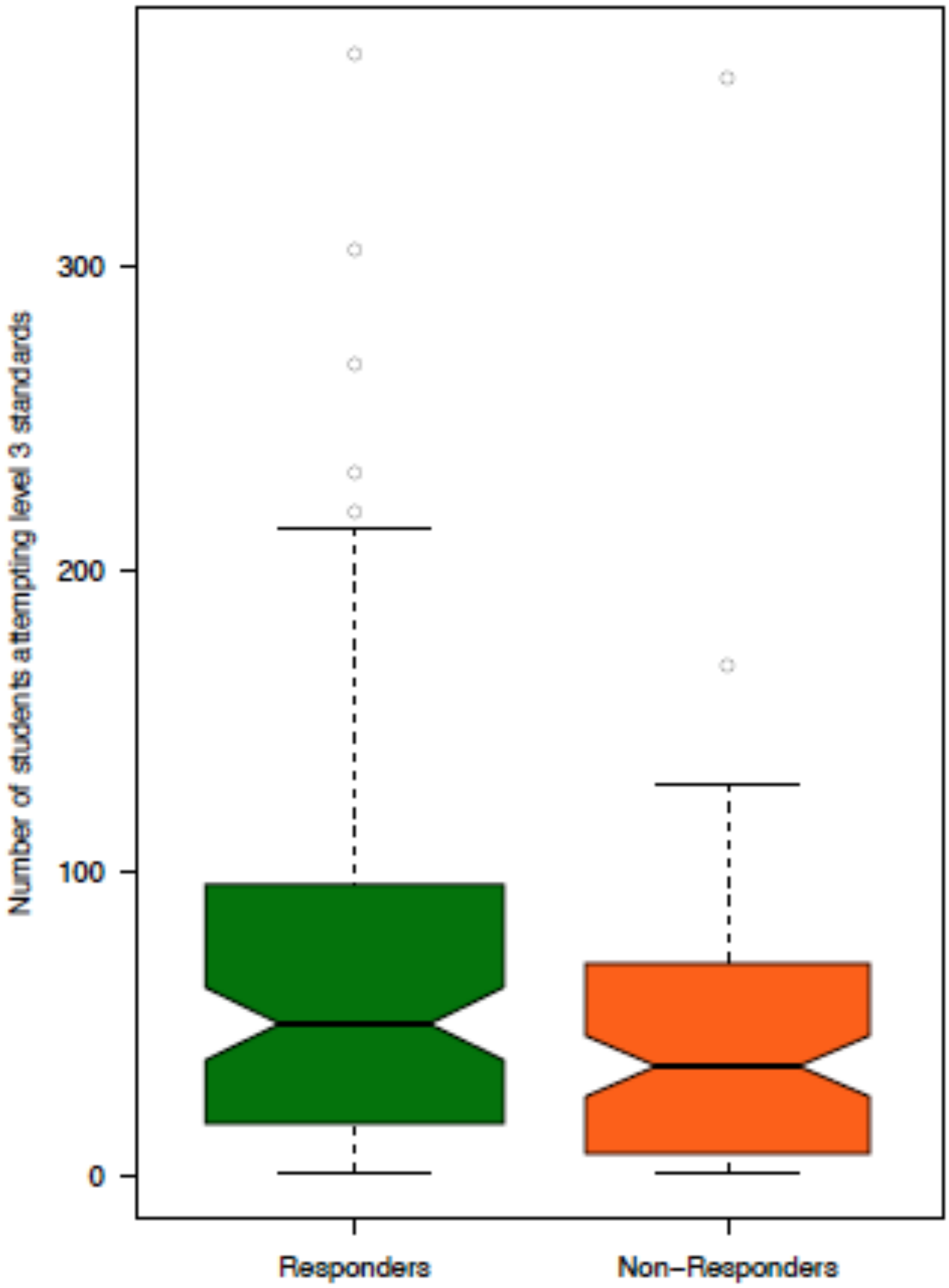

Figure 22: Box-and-whisker plots of a measure of school size, comparing responders and non-responders

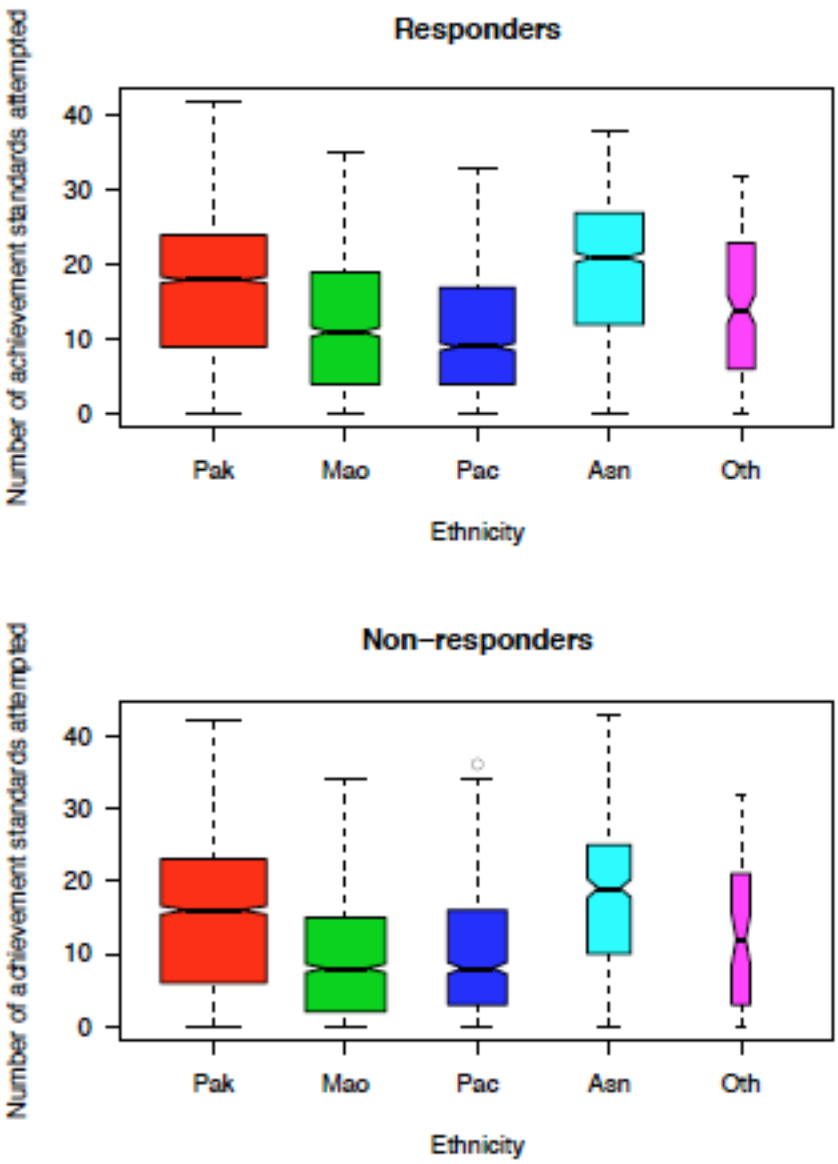

Figure 23: Box-and-whisker plots of the number of achievement standards attempted, by ethnicity, comparing responders and non-responders

Note: The widths of the boxes in these box-and-whisker plots are proportional to the square roots of the corresponding sample sizes.

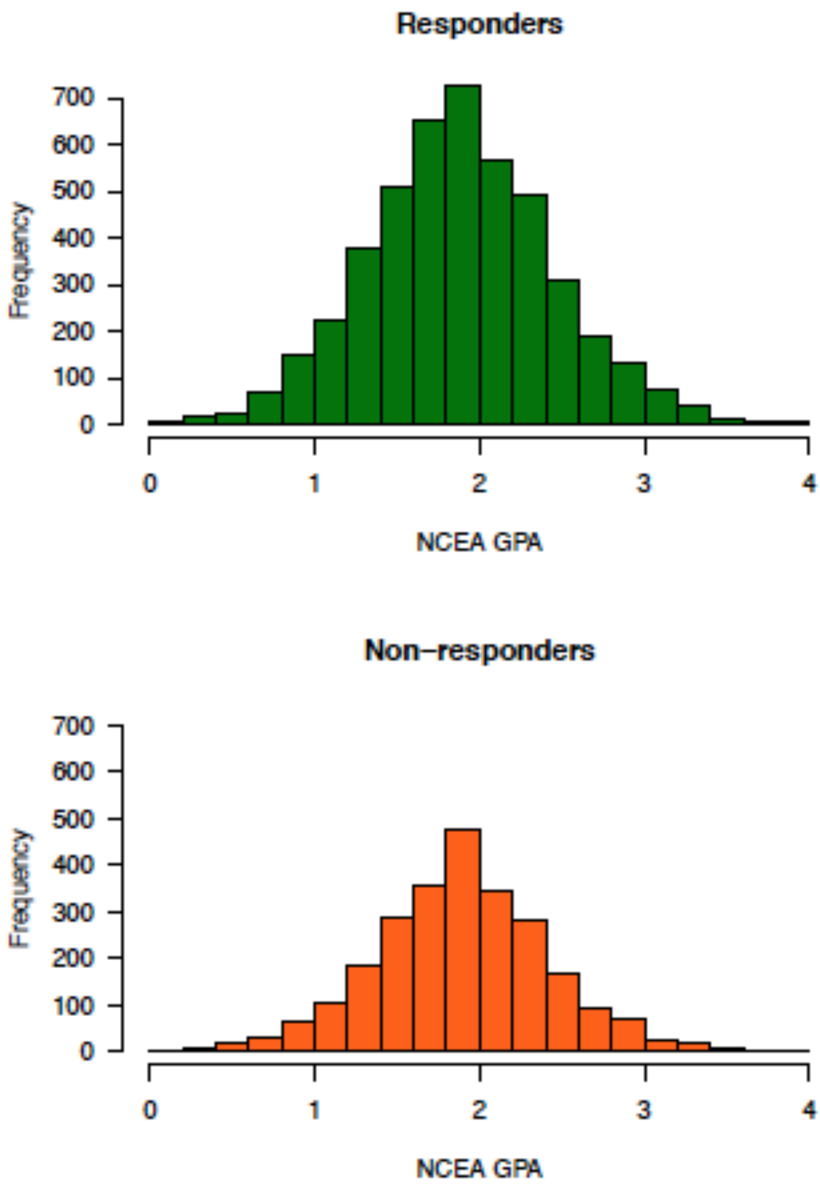

Figure 24: Histograms of NCEA GPA, comparing responders and non-responders

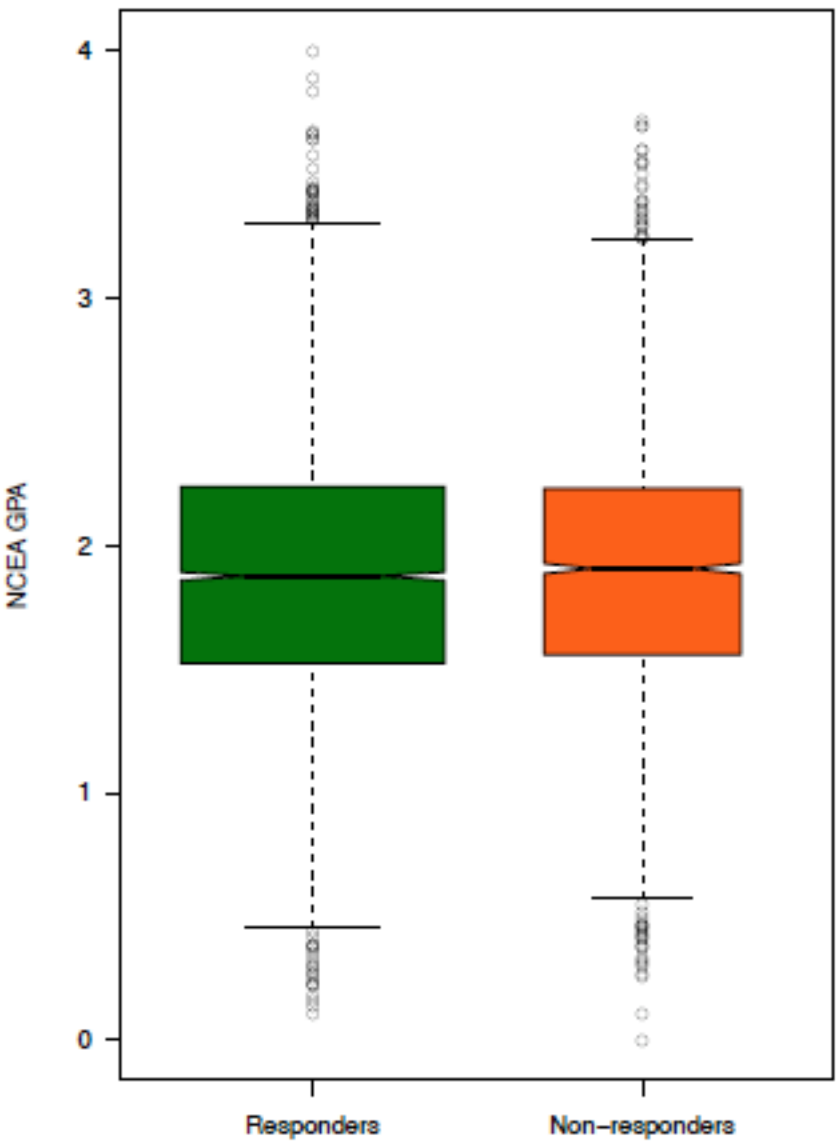

Figure 25: Box-and-whisker plots of NCEA GPA, comparing responders and non-responders

Note: The widths of the boxes in these box-and-whisker plots are proportional to the square roots of the corresponding sample sizes.

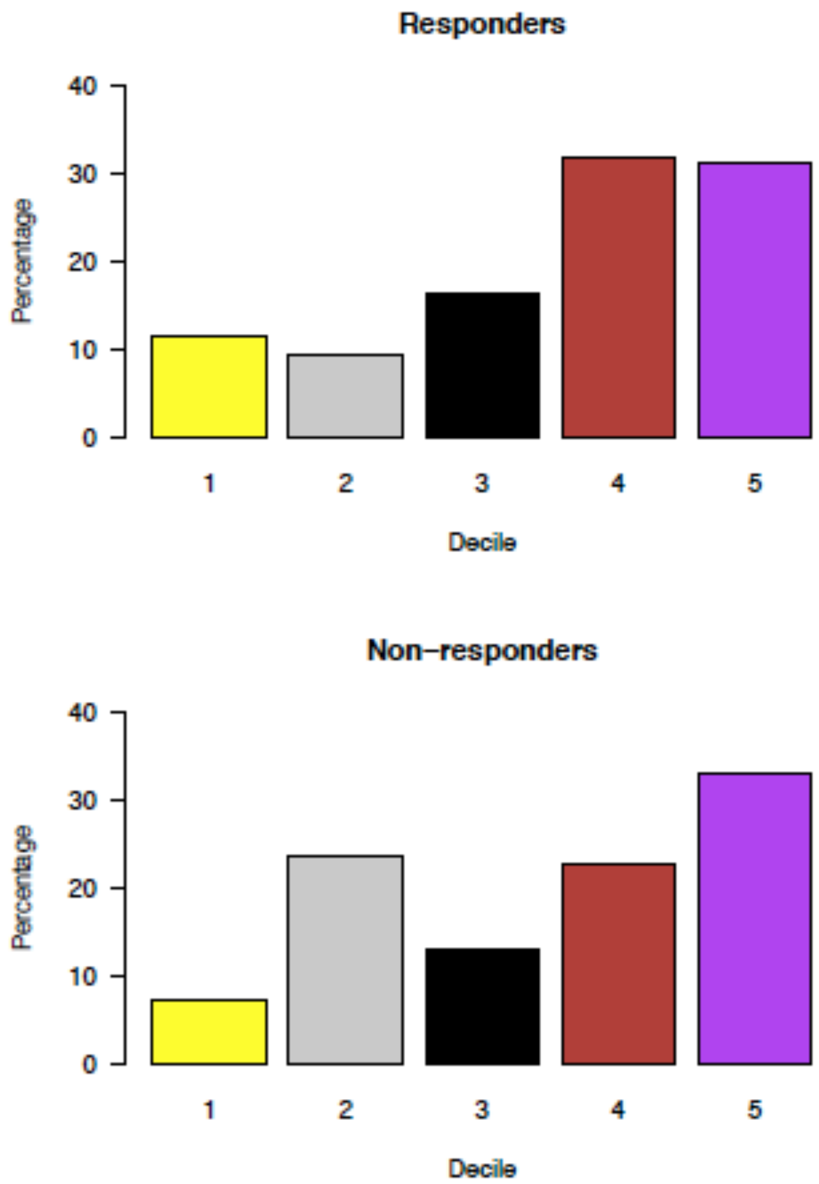

Figure 26: Bar plots of decile counts, comparing responders and non-responders

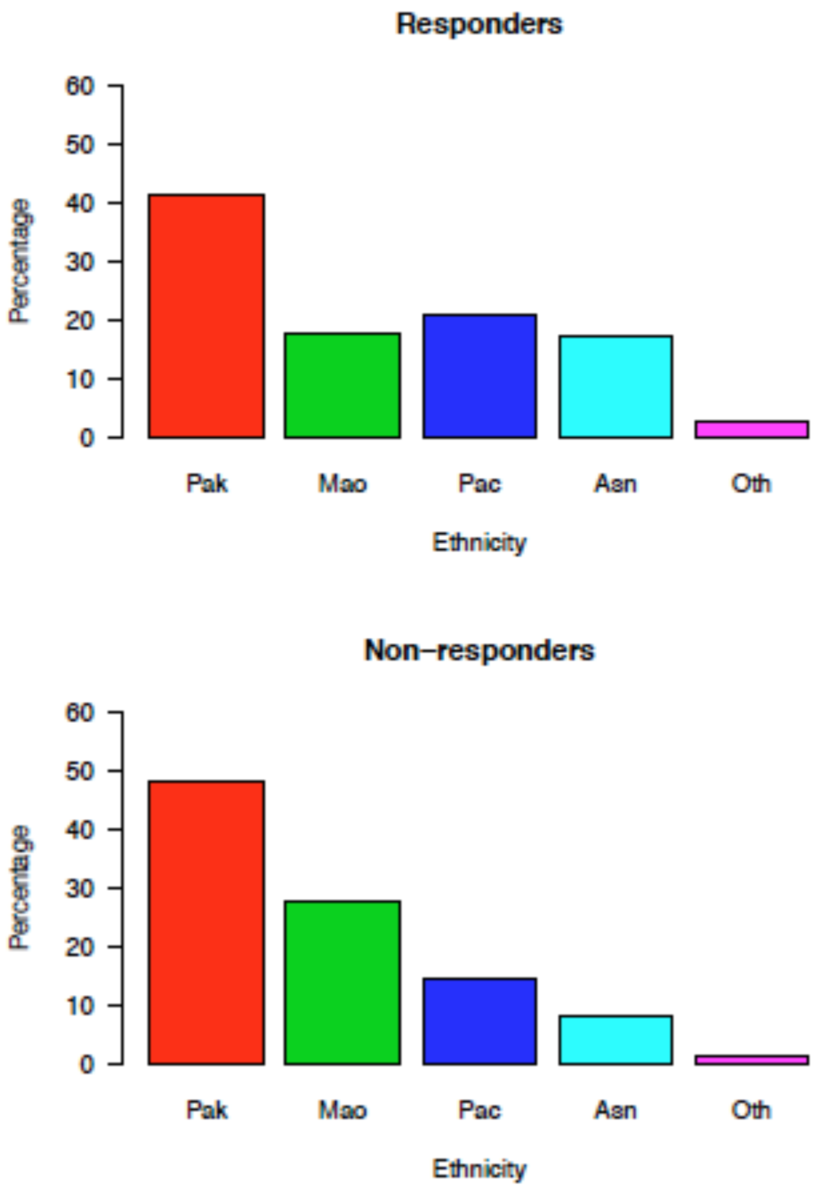

Figure 27: Bar plots of ethnicity counts, comparing responders and non-responders

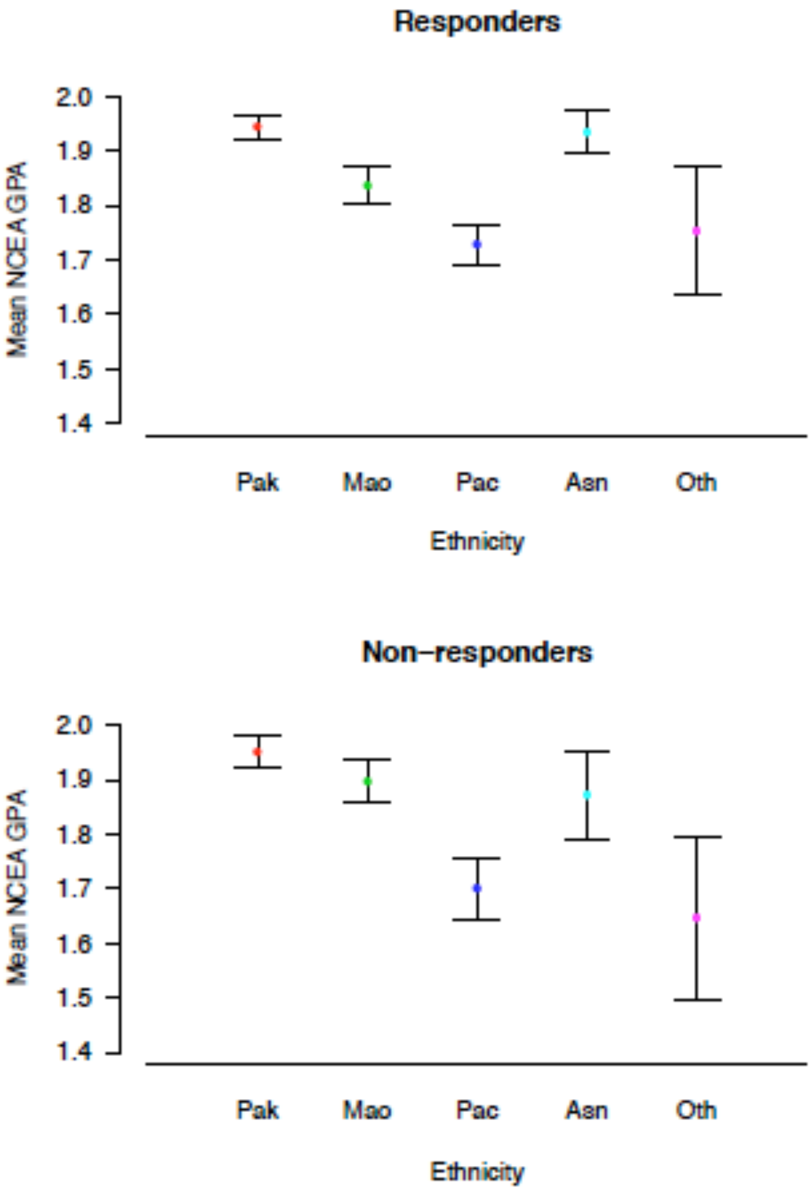

Figure 28: Mean NCEA GPA, with 95% confidence error bars, by ethnicity, comparing responders and non-responders

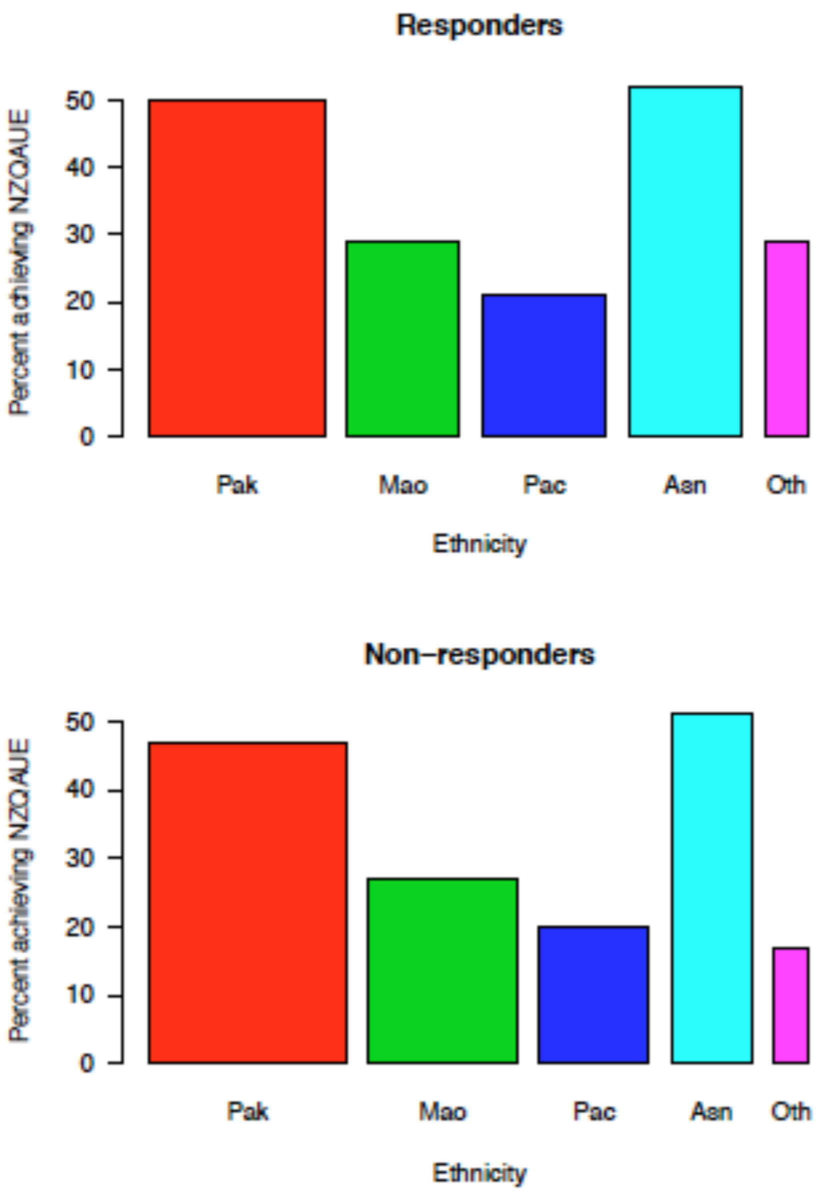

Figure 29: Bar plots of percentages achieving NZQA UE by ethnicity, comparing responders and non-responders

Note: The widths of the bars in these bar plots are proportional to the square roots of the corresponding sample sizes.

A University of Auckland Partnership for Excellence

Starpath Project
The University of Auckland,
Epsom Campus, Faculty of Education
Private Bag 92019, Auckland
starpath@auckland.ac.nz
www.starpath.auckland.ac.nz