

# **The Importance of Leaving School Before the Exam in Analysing Academic Performance\***

**Guyonne Kalb**  
**Principal Research Fellow**  
**The University of Melbourne**

**and**

**Sholeh A. Maani**  
**Associate Professor**  
**The University of Auckland**

## **ABSTRACT**

In this paper, we use a recent panel data set from New Zealand to examine the link between the academic performance and the decision of teenagers to drop out of school before the exam at the end of year 10. These choices have significant lifetime economic impacts, since early school leaving in many cases closes pathways to further education. We address the potential endogeneity and error correlation of potential performance in national examinations and earlier school leaving choices. Birth month provides an instrument to be used in the equation for drop out, because those born in particular months can legally leave school before the exam takes place whereas the other students cannot do this. The analyses incorporate the effect of academic ability (childhood IQ and scholastic measures), parental education, family resources at different points in time while the child is growing up, and school and peer characteristics. The results show that those who drop out early are unlikely to have performed well in the exam. In addition, the difference between those who drop out and those who continue at least up to their exam is almost completely explained by observed factors. Leaving out those variables which are often not available in other datasets, such as childhood IQ or childhood family resources information, we find that the unobserved factors in academic performance and early school leaving are correlated.

**JEL classifications: I21, J13, J24, J18**

---

\* This study was made possible through a research grant from the New Zealand Treasury. Special thanks go to Drs David Fergusson and John Horwood for making the data set available and for insightful information on the data set, Drs Robert Haveman, Barbara Wolfe, Stephen Jenkins, Lawrence Kahn, John Ermisch for insightful comments on the school leaving question, Drs George Barker, and Ron Crawford, for helpful comments on an earlier part of the study, and Adam Warner for initial research assistance with the data set. None of the above is responsible for the views expressed.

## I. Introduction

A growing body of economic research is focussing on the Academic Performance of children and adolescents, as an important economic indicator of investments in education by families and communities. Almost all academic performance data, such as SAT scores (Scholastic Aptitude Tests) for College Entrance, have a self-selection aspect, as to the personal choice of taking the examination. A question of interest is whether this selection feature of academic performance data, due to either the choice of taking the examinations or dropping out of school early, is expected to affect the results in estimations of academic performance.

This paper uses a recent panel data set from New Zealand to examine the link between academic performance in national examinations and previous school-leaving choices by adolescents. A model of academic performance, allowing for selection through school-leaving choices, is estimated and potential error correlation between performance and school-leaving choices is addressed. The analysis incorporates the effect of academic ability (IQ and other childhood scholastic measures), parental education, family resources over the child's growing years, and school and peer characteristics. The estimated model can be used to make inferences on the academic performance of early school leavers, even though no examination results are available for this group.

The paper addresses three questions. First, whether the academic performance of those taking the exam is representative of the complete student population including those who left school early. Second, it compares the predicted academic performance of those leaving school early versus the predicted academic performance of those taking the examination. Finally, we examine the effect of excluding those variables that are often not observed in other data sets from the model. The excluded variables include childhood IQ, early childhood family resources and peer effects. This allows us to examine the results when only using the information commonly available in data sets

Demand for post-compulsory education as a personal choice has been addressed by Willis and Rosen (1979), who estimate participation in university studies, and by Rice (1987), who estimates secondary school leaving in Britain. Early school leaving, at the post-compulsory level, is observed internationally. The question of the effect of parental resources on academic performance of children and adolescents and their later schooling outcomes has, in turn, received recent attention by for example, Blau (1999), Feinstein and Symons (1999), Ermisch and Francesconi (2001), Maani (2006), and Maani and Kalb (forthcoming, 2007). This study extends the literature by estimating models of school leaving jointly with models of academic performance and addressing the above three questions.

Similar to many other countries, education is compulsory in New Zealand up to age 16. In addition, those who are at school at age 15 are expected to take the School Certificate Exams at the end of their Year 10 class. These exams, which are nationally administered, are based on the same set of questions and grading for all participants. This is a great advantage as the use of such a measure of academic performance eliminates problems with the potential inconsistency in comparing grades across schools in lower and higher income decile localities. It thus provides nationally comparable academic performance results at age 15 and prior to the school leaving choice once students turn 16. This differs from, for example, SAT scores in the US, which are given at the end of high school. However, a small group of students starts school later and would turn 16 before the exams. Therefore, they could leave school legally before the exams.

The innovations of this study are the estimation approach, the addressing of the link between academic performance and early school-leaving choices of young adults, and the use of longitudinal data to explain schooling outcomes. This is made possible by the use of the Christchurch Health and Development Study (CHDS) data set<sup>1</sup>, which allows us to control for important longitudinal personal and household conditions starting from birth. Examples of variables are annual parental income decile from early childhood to teenage years, the child's IQ at age 8, and a measure of academic performance at age 15, which is standardised through a national set of examinations. These features of the analysis substantially reduce the 'unexplained' part of estimations due to heterogeneity, and make it possible to control for both the set of factors that are related to personal academic *ability*, and childhood and adolescent parental *resources*. The ability to make a distinction between early childhood and teenage household income conditions is of interest in examining the separate effects of parental income at different points in time on academic performance and early school-leaving choices.

The plan of the paper is as follows. A brief presentation of the analytical framework is provided in Section II. A discussion of the data set and the characteristics of the sample follow in Section III. The estimated models and their results are presented in Section IV, followed by concluding remarks in Section V.

## II. Analytical Framework and Estimation

The theoretical modelling framework, which is widely adopted in the economic literature on participation in post-compulsory education, focuses on individual choice for long-term investment in human capital and the inter-temporal nature of the investment decision (e.g. Becker, 1993; Schultz, 1961).<sup>2</sup>

The decision to participate in higher education and training is intrinsically related to a number of factors. For example, investment in higher education is expected to result in higher returns for those with greater ability and a taste for lifetime labour force participation. In addition, household financial constraints would influence the cost of obtaining education. Moreover, the family socio-economic background can affect the demand for post-compulsory and higher education through tastes, and the costs of obtaining information.

Therefore, *ceteris paribus* those individuals who have higher academic ability and a stronger taste for earned income as opposed to leisure over their lifetime are more likely to invest in higher education. In addition, keeping ability constant, a greater potential to finance education will lead to greater participation. The model can further control for other personal characteristics such as age and gender.

An extended framework for analysing participation in higher education is based on the Willis and Rosen model (1979) of participation in university studies in the U.S. and Rice's (1987) extension to secondary school leaving in Britain. In this framework, choosing a level of education depends on the expected value of lifetime earnings at that education level, and also on background characteristics, which determine the individual's tastes, expectations, and on the

---

<sup>1</sup> For further information and other research using this data set the reader may refer to Fergusson et al. (1989), Fergusson et al. (1991), and Fergusson and Lynskey (1993).

<sup>2</sup> A detailed discussion of the human capital framework for the study of participation in higher education, and a review of the empirical literature on participation in post-compulsory and higher education is provided in Maani (1996 and 1997).

financial constraints facing the household. Individuals select different levels of education on the basis of financial resources, tastes, perceptions and natural ability. Therefore, individuals are sorted into education levels (for example, school leaving at age 16 or continuation to post-compulsory levels) according to an underlying joint distribution of tastes, talents, expectations, and parental income. These characteristics are assumed to be randomly and independently distributed across individuals. While Willis and Rosen's analysis utilised structural models and emphasised self-selection, Rice's application utilised reduced form models of participation and emphasised the effect of financial constraints on school leaving choices of males and females.<sup>3</sup> Neither study had observable variables on academic ability such as IQ or academic test scores.

In this framework, educational choices for the  $i$ th individual are influenced by  $V_{ij}$ ,

$$V_{ij} = V(E_j(S_i), X_i, u_i), \quad j=0,1. \quad (1)$$

where  $V_{ij}$  is the utility of net expected present value of life-time earnings at each level of educational attainment  $j$  ( $E_j$ ), as influenced by individual talents and abilities ( $S_i$ ); and observable personal and environmental characteristics ( $X_i$ ), which determine the individual's tastes, expectations and the financial constraints facing the household, and  $u_i$  are the unobservables. Thus, the individual invests in additional education beyond the compulsory level if the expected net benefits are positive ( $V_{i1} - V_{i0} = G(S_i, X_i, u_i) > 0$ ).

Empirical estimation of the probability of enrolment at post-compulsory education (Pr PCE) is based on equation (2) below:

$$\text{Pr PCE observed} = \text{Pr} [ (V_{i1} - V_{i0} = G(S_i, X_i, u_i) > 0 ) \quad (2)$$

where vectors of observables  $S_i$  and  $X_i$  result in observation of participation if  $V_{i1} - V_{i0}$  is positive, and  $u_i$  are unobservable characteristics. Given the assumption that the distribution of net benefits conditional on  $S_i$  and  $X_i$  and their underlying characteristics are normally distributed, Pr PCE would also follow the standard normal cumulative density function and can be estimated via Probit analysis, such that

$$V_{i1} - V_{i0} \sim N ( S_i' \beta + X_i' \gamma , \sigma^2 ) \quad (3)$$

with  $\beta$ ,  $\gamma$  and  $\sigma^2$  constant across the population (e.g. Willis and Rosen, 1979; Rice, 1987).<sup>4</sup>

<sup>3</sup> It is interesting to note that although the Willis and Rosen (1979) model is based on Human Capital theory, it is also consistent with Signalling theories of investment in education, since in both theories schooling is pursued to the point where its marginal (private) internal rate of return equals the rate of interest. Both theories are also consistent with the model in which participation in education is influenced by the capacity to finance education, ability, tastes, perceptions and information, and expectations (some observed and some unobserved). However, in human capital theory, investment in education is assumed to increase labour productivity, while in signalling theory, education is a positional good to signal information on unobserved ability.

<sup>4</sup> The above model is nested in a model of lifetime utility maximisation, which determines labour supply and education investment decisions. Although it is possible to estimate empirical models, which are based on joint determination of expected future labour supply and participation in higher education if sufficient data is available, the education participation model above presents a satisfactory approach by providing a reduced-form model of participation, which incorporates the effect of tastes and ability. In addition, the lifetime supply decisions of young persons have not materialised at the time of participation in education. These can at best be measured empirically as stated expectations or intended behaviour influenced by the same set of factors that determine the participation in

Academic performance, as measured by the average score in a set of examinations, can in turn be modelled as:

$$A_i = f(S_i, X_i, v_i) \quad (4)$$

where  $A_i$  is the score of individual  $i$  (representing ability and effort);  $S_i$  represents personal talents and abilities; and  $X_i$  is a vector of personal and parental resources, and environment; and  $v_i$  represents the effect of unobserved factors, such as motivation.

Academic performance of children and adolescents has recently received significant interest in the literature in models that link parental resources to children's academic performance.<sup>5</sup> The recent literature on educational attainment has emphasized the significance of parental investments in human capital since childhood. An important implication of this literature is the recognition that even if teenagers have a significant input into the decision of when to terminate secondary schooling, they are constrained in their choice by academic performance and human capital investments throughout childhood (Ermisch and Francesconi, 2001).<sup>6</sup> The models of *children's* academic performance are usually based on a production function, where the parents are the producers. However, the academic performance of teenagers at the end of year 10 and their school-leaving choices around age 16 are likely to be joint decisions, influenced, amongst other things, by the adolescent's personal ability and human capital investments by parents throughout childhood.

If  $A_i$  were a continuous measure, equation 4 could be estimated through Ordinary Least Square estimation. A Tobit approach is used in this case, because of the censored nature of exam grades below the Fail and above the grade A cut offs. That is, the academic performance of the most capable and the least capable cannot be accurately measured through an exam targeted at the average student.

In the discussion of the modelling approach in Section IV, we consider joint estimation issues of equations 4 and 2, and deal with error correlation across academic performance and school-leaving choices of individuals. This accounts for potential self-selection into school leaving, which may be related to (expected) academic performance.

Our study further extends earlier studies by examining the effect of personal ability and parental resources at different points in time on both academic performance and school-termination choices of young adults, using a data set, which includes a remarkably large number of relevant variables, thereby reducing the importance of the unobserved components. Notwithstanding the richness of the data, the joint estimation of the equations for academic performance and dropout from school before the exam allows us to formally examine and account for unobserved factors influencing both school leaving and academic performance.

---

education decisions.

<sup>5</sup> Blau (1999), Borjas (1995), Case and Katz (1991), Duncan et al. (1998) and Montgomery (1991) for the US; Feinstein and Symons (1999) and Ermisch and Francesconi (2001) for the UK; and Miller and Volker (1989), Prior and Beggs (1989), and Borland and Wilkins (1996) for Australia are examples of studies on the link between parental resources and educational attainment or labour market outcomes for their children.

<sup>6</sup> Ermisch and Francesconi (2001) consider seven levels of educational attainment in the UK, from no qualifications to degree qualifications, and estimate ordered logit models of educational attainment. Other studies such as Feinstein and Symons (1999) have focused on test scores in maths and English at age 16 in the UK.

### III. Characteristics of the Sample

The Christchurch Health and Development longitudinal Study (CHDS) includes extensive economic and academic information on a cohort born in Christchurch in 1977. This cohort is followed throughout their childhood and adolescence, providing information on their transition from school to further education, training and work. Among the advantages of this data set is the extensive amount of information on the cohort's academic and home environments, academic performance and ability, and socio-economic background.

The sample analysed in the study utilises information from survey years from birth in 1977 to age 16 of the cohort, selecting respondents for whom data on all variables of interest was available.<sup>7</sup> The characteristics of this sample are summarised in Table 1 below.

As column 1 of Table 1 on the full sample shows, about half of the sample (50.5%) was female. The characteristics of the sample on academic performance and economic conditions are reassuring in relation to expected national averages, such as the average IQ of 102.8, and the average school certificate mark of 1.06 or a C. Home ownership by parents was 88.6%, and the average proportion of family income from benefits was 13.9%. In the sample, 7.4% were Maori and 2.8% were Pacific Islanders.

#### [ Table 1 about here ]

In relation to parental education, 49.8% of the mothers and 47.5% of the fathers of the respondents had no school qualifications (less than the year 10 School Certificate), and 20.6% of mothers and 19.5% of fathers had tertiary qualifications.

Column 2 in Table 1 presents the mean characteristics for those who had not dropped out before exams. These characteristics include the individual's IQ at age 8, the average School Certificate grade obtained (reflecting academic factors), the household income decile between ages 11 and 14, and in early childhood between ages 1 and 5, and school, neighbourhood and peer factors, such as the proportion of the student's class continuing to post-compulsory levels (at age 16), or association with peer groups with deviant behaviour (a 1-10 scale reflecting problems with the law, substance abuse, etc.).

In general, mean group characteristic comparisons, as reflected in formal modelling in the next section, show that those who participated in post-compulsory secondary schooling at age 16 had mean characteristics, which were different from those of school leavers. These differences included a higher average IQ at age 8, higher average School Certificate marks, they belonged to a higher family income decile, and they went to a school with a higher proportion of the class

---

<sup>7</sup> The original cohort of individuals in the survey consisted of 1265 individuals. The sample used in this study contains 713 observations to analyse the dropout before exams and 598 observations for the joint estimations of School Certificate Examination at age 15 and school leaving at age 16. The smaller sample used for age 16 is partly due to minor attrition over time, and partly due to missing values on variables of importance to this part of the study, such as IQ, parental income, and school factors. Analysis indicates that the selected sample is slightly less likely to drop out of secondary school than the full sample (the probability is 0.0034 lower). A study for the New Zealand Treasury (Maloney, 1999) showed that attrition was related to some initial characteristics such as ethnicity and having a single parent. However, comparisons with later Census data at both local national levels show that the CHDS is still fairly representative of families with children born around 1977.

continuing to the Sixth Form. These characteristics are consistent with the hypothesis that individuals sort themselves into different choices based on their academic ability and the expected returns to their choice, family income constraints, and influences from their school and peer environment. They are further consistent with the observation that adolescents from the lower income deciles are more likely to leave school early and less likely to participate in higher studies.

## IV. Models and Results

In this section, the issue of school leaving before the exams is analysed to explore the expected academic performance of students who leave before taking the exam. We compare these expected results to the results of those taking the exam. For the students leaving before the exam no measure of academic performance is available, but fortunately, information on all the other factors is available. This means we can estimate an academic performance equation representing the whole student population by accounting for the selectivity of school leaving before the exam through joint estimation of academic performance and leaving before the exam, accounting for correlation between unobserved factors in both equations. In the first subsection, school leaving before the exam is discussed briefly. The second subsection discusses the academic performance equation and combines it with the school leaving equation. The model is first estimated with all available explanatory variables, and then with a selection of the explanatory variables, to investigate the sensitivity of the results to including only those variables that are usually available in datasets.

### IV.1 School Leaving

This model examines the effect of personal characteristics such as cognitive ability and academic performance, peer and school effects, and parental economic constraints on the choice to drop out of school at age 16 and before the national exam. The dependent variable is binary as to whether or not the respondent had left school before the exam, as opposed to enrolment until after the exam. The sample of analysis excludes: those who do not leave school before the exam and have missing information on the average score; or have zero subjects in which they were enrolled for the exam; and those who have missing information on any of the included explanatory variables. We use a Probit model to analyse school leaving. Explanatory variables  $X_i$  represent personal characteristics, such as ability, gender, socio-economic and cultural background. They further represent household and environmental constraints such as household assets, proportion of household income from government welfare benefits, school effects, and neighbourhood and peer effects.<sup>8</sup>

The measures of parental income we have included reflect permanent rather than current income providing a measure of long-term parental resources (see, Blau (1999) for US evidence based on the NLSY data set). In addition, the data allow us to distinguish between parental income effects during early childhood and adolescent years on later academic performance and school leaving choices. The two income measures are correlated, but the correlation is only 0.55 and thus each measure provides some independent information on the financial history of the household. Duncan et al. (1998), for example, find evidence for the US, based on the PSID data set, that family economic conditions in *early childhood* are more pronounced determinants of completed schooling years than economic conditions later in life.

---

<sup>8</sup> See Table A1 in Appendix A for a full list of variables and their descriptions

The proportion of income from benefits was calculated based on data on all sources of parental welfare benefit income and other sources of income. The variable reflects the relative significance of benefit income compared to the young person's family income. The variable also reflects beneficiary status and relative disadvantage with regard to the household's wealth and assets.<sup>9</sup>

## IV.2 Academic Performance and School Leaving before the Exam

The dependent variable in the academic performance analysis is the average National School Certificate Examination grade, which is normally taken at age 15, on five subjects. The average is a continuous variable, which ranges in value between 3 for an A average to 0 for a D (fail) average.

The average grade on all subjects is censored at 0 and 3. The score D for a fail is translated into a value of 0, a score C into 1, a score B into 2 and a score A into 3.<sup>10</sup> The dependent variable is constructed by averaging the numeric values of the score over all subjects taken in the certificate. Thus, the minimum score is 0 and the maximum score is 3. Individuals below and above a certain academic performance level cannot be ranked besides observing that they are at the minimum or the maximum level.

The model for latent academic performance  $A_i^*$  looks as follows:

$$A_i^* = \alpha_a + X_{ai}'\gamma_a + \varepsilon_a \quad (5)$$

However, we only observe  $A_i$ , the average score, which is censored at the lower and upper end:

$$A_i = A_i^* \text{ if } 0 < A_i^* < 3 \quad (6)$$

$$A_i = 0 \text{ if } A_i^* \leq 0 \quad (7)$$

$$A_i = 3 \text{ if } A_i^* \geq 3 \quad (8)$$

This equation is estimated using a standard Tobit model.

We can only observe the average score for people who stayed on at school to the time of the exam. Therefore, a first step in exploring the results from an academic performance (average grade) equation is to estimate a Tobit regression allowing for selectivity of students leaving school before the exams. That is, we estimate a joint model for school leaving and academic performance. The equation for school leaving contains one variable not included in the academic performance equation, which is birth month. Children in the survey are born between April and August. Those born in April or May start school in the first semester of the year they turn five and take the first kindergarten equivalent year only once. However, those born in June to August start school in the second semester of the year, when they turn five, and usually have a full kindergarten year in the following year. The second group will turn 16 before the national exams, allowing them to leave school legally before the exams are administered. Children born

<sup>9</sup> Rice (1987) used a 'current income' variable in addition to the 'benefit ratio' (the ratio of current benefit to current household income). In this study, the definition of the income and benefit variables is different from the Rice study, in that income is measured as the average family income decile between the ages of 11 to 14 and 1 to 5. Since the benefit ratio in this study is the only measure of current income (that is, household income when the child is aged 16), it would explain why benefit ratio could have a negative effect on school retention, partly reflecting the effect of economic disadvantage.

<sup>10</sup> This is also compatible with the official GPA (Grade Point Average) score assignment in New Zealand.

in the earlier months of the year only turn 16 after the exams. Therefore, birth month is expected to affect school leaving but not academic performance, and it is used to identify the model in this paper.

#### **IV.2.1 Results Using All Variables Available in the Data Set**

Academic performance at the end of grade 10 is expected to be influenced by many of personal, school and family resource variables, which influence school-leaving choices at age 16 as well. That is, academic performance is influenced by parental education and resources, personal ability, and school and peer effects, and these are all factors which are at work over a number of years. In addition, the same set of unobserved variables can potentially influence both academic performance and school leaving. With the use of this dataset, we are able to control for a remarkably large number of important variables, which are often not observed in other data sets, reducing the error component of the two equations. Table 2 presents the results for the joint model.

[ **Table 2 about here** ]

The academic performance equation (the top section of Table 2) shows that academic performance is influenced by a host of personal and family factors since childhood, which have been at work for a long period of time. Childhood IQ, parental income in the past, parent's education, and school and peer effects are all important explanatory variables in determining academic performance. In addition, girls perform better academically.

Both early childhood and more recent parental income matter. Including both recent and early childhood income decile variables, we find that both early childhood and adolescent parental income deciles are important in explaining academic performance. The effect of each higher recent income decile (averaged over the time when the respondent was aged between 11 and 14) is estimated to be equivalent to 0.060 of a grade in the exam. In addition to this effect, early childhood income decile explains an additional effect of 0.039 of a grade. Therefore, keeping other factors constant, together the predicted effect of the income decile variables is close to 0.1 of a grade difference for each income decile or close to a complete grade for 10 deciles difference (the difference between a C or a D average grade, for example). These results are consistent with the effect of poverty (Duncan et al., 1998), and the positive effect of income in many US studies (Haveman and Wolfe, 1995), and Gregg and Machin's (1998) findings on the effect of financial difficulties in early or late childhood. However, the finding by Duncan et al. (1998), that only early childhood parental income is significant in explaining the years of completed schooling and high school completion, is not repeated here. These results are further consistent with recent results for the UK (Feinstein and Symons, 1999; and Ermisch and Francesconi, 2001) regarding the importance of resources throughout childhood in determining children's academic performance.

While the correlation between early childhood and the later income decile variable was 0.55, a large number of young adults in the sample had experienced changes in their family's income decile, between their early childhood and adolescent years. Some respondents experienced improvements in their household's relative income position, while others experienced deteriorations.

In addition, the mother's lack of school qualifications, and class and peer effects are significant in explaining academic performance. Personal academic ability is also important. Each additional unit of IQ score is associated with 0.030 of a grade. This is a large effect considering

the range of IQ scores. The mean IQ score was 102.8 with a standard deviation of 14.8, a minimum of 70 and a maximum of 143. This effect highlights the importance of the respondent's childhood scholastic ability in predicting academic performance in later years. Comparing this effect to the combined effect of the two income decile variables, we see that one income decile is equivalent to slightly more than 3 IQ units.

The results of the school-leaving Probit indicate that, keeping other factors constant, children of parents without qualifications and children who associated with deviant peers are less likely to stay at school until after the exam. Children of parents who own their home, children who experienced higher childhood family income and children who had a higher IQ at age eight are more likely to stay at school. Once IQ and peer effects are included in the model, relatively few other variables are significant. For example, the variables for ethnic background and most indicators for parental education were insignificant. The identification variable birth month is significant at the 10 per cent level and has the expected sign. Those who cannot leave school legally before the exam are more likely to stay at school until after the exam.

The correlation coefficient  $\rho$  is small positive and insignificant, indicating that self-selection into taking the exams as a result of unobserved factors is not a major issue. Comparing the academic performance equation in the joint model with the equation for staying at school until the exam to a single-equation model for academic performance, it is clear that the results in both models are very similar. This is as expected, given the small insignificant value for the correlation coefficient.

From the results discussed above, it is interesting to note that once personal, socio-economic and environmental characteristics are controlled for, Maori and Pacific Island teenagers do not perform more poorly and they do not have a statistically significantly higher probability of leaving school before the exam.

Similar results are obtained when using Ordinary Least Squares for the academic performance equation combined with the same selection equation as in Table 2, using a two-step Heckman approach.<sup>11</sup> The alternative results are presented in Table 3 and are similar to the results in Table 2. Using the two-step approach shows that the inclusion of a Heckman adjustment term in the academic performance equation results in an insignificant selectivity coefficient and small changes in the other parameters.

[ Table 3 about here ]

---

<sup>11</sup> In this approach, the selection equation is estimated separately and used to construct a correction term controlling for the selectivity of students taking the exams. A Probit equation is estimated for the probability of staying at school until after the exam:  $\Phi^{-1}(P_{di}) = Z_i' \gamma_Z$

(9)

The estimated parameters from this equation are then used to construct a Heckman correction term, so that the results can be extrapolated to the whole population including those who dropped out. This is added to the OLS equation for academic performance. This allows for the selectivity in the model of average scores so that we can predict average scores for people who took and those who did not take the exams. The academic performance equation is then:

$$A_i^* = \alpha_a + X_{ai}' \gamma_a + \lambda_i \gamma_\lambda + \varepsilon_a \quad (10)$$

where  $\lambda_i = -\frac{\varphi(Z_i' \gamma_Z)}{1 - \Phi(Z_i' \gamma_Z)}$  is the Heckman correction term for those who took the exam

$\phi$  is the standard normal probability density function and  $\Phi$  is the standard normal cumulative distribution function.

To explore the implications of the estimated models with and without allowing for correlation in unobserved heterogeneity between school leaving and academic performance, we use the estimated parameters in Tables 2 and 3 to predict the average grade for everyone in the sample using the different models.<sup>12</sup> Independent of the approach taken, we find that those who had dropped out before the national exams were expected to perform poorly and significantly below those who did take the national exams (see Table 4). This lower predicted score was mostly due to a difference in observable characteristics between the two groups. The first row for each category in Table 4 provides a predicted average grade (Ave\_Grade\*) based on the model controlling for sample selection from Table 2, whereas the second row provides a prediction without the control for sample selection (Ave\_Grade), the third and fourth row provide similar results based on the OLS results from Table 3. The predictions in the first rows, for example, show that those who took the national exams had a predicted average grade of C (mean = 1.1751) compared to a predicted grade of D (mean = 0.2468) for those who did not. Controlling for sample selection makes the results more pronounced by predicting lower average grades close to zero for those who did not take the national exams and dropped out of school by age 16. However, the results from the different approaches were quite similar. This result implies that those who had dropped out of school before taking the national exams were expected to perform poorly in the exams, and importantly this seems mostly caused by a difference in observed characteristics rather than through unobservable characteristics.

[ Table 4 about here ]

#### IV.2.2 Sensitivity of Results to Excluding Often Unobserved Variables from the Analysis

In this subsection, we analyse the effect of leaving out important variables from the model. Many datasets are less rich with regard to the range of variables available in them than the New Zealand data used in this study. We are interested in the effect this would have on the model's ability to distinguish the academic performance of school leavers from those staying at school until after the exam. For this reason, we re-estimate the models in Tables 2 and 3, leaving out the important but usually unavailable variables of childhood IQ, early childhood family resources, association with deviant peers at age 15 and the proportion of the individual's 10<sup>th</sup> grade who go on to the 11<sup>th</sup> grade. The same sample as before is used for this analysis as for the analysis in IV.2.1, even though more observations would be available due to having fewer non-missing explanatory variables. This allows for a straightforward comparison of the results: any differences must be due to leaving out these four characteristics.

The results are presented in Tables 5 and 6. Both tables show a clear difference in results, with the correlation now being more significant in the Tobit specification and larger in size (in absolute terms) in the OLS. The correlation coefficient now has a negative sign in both specifications. Other differences are that the effect of parental education and being from Pacific Islander descent are now more significant, as is the effect from income decile. The constant term in the academic performance equation now has a small positive value instead of a negative value to compensate for excluding the positive effect from childhood IQ, where higher IQ levels would result in higher academic performance. In this limited-variable specification of the model, the constant incorporates an average academic performance across all IQ levels. Nevertheless, the

---

<sup>12</sup> Using the estimated parameters from the dropout equation ( $\gamma_z$ ), an estimate of  $\lambda$  is constructed and included in the academic performance model. Average scores for those who dropped out before the exams can be predicted using

equation (10). For this group, the correction term is  $\lambda_i = \frac{\varphi(Z_i' \gamma_z)}{\Phi(Z_i' \gamma_z)}$ .

correlation term is still not significant and has a sign opposite from what you would expect if you leave out a variable that is positively correlated with both dependent variables.

Table 7 presents the predicted average scores based on the models in Tables 5 and 6. Again, the results are clearly different from those presented in Table 4. As before, the differences between models that account for potential self-selection into taking the examination and the models that do not account for this are quite small. This is as expected given the relatively low and insignificant value for the correlation parameter. However, the largest difference is found for the predicted academic performance of those who were not taking the exam. Although the specification that does not allow for self-selection into taking the exam could not be rejected, it is clear much less has been captured by the observed heterogeneity than in the full model.<sup>13</sup> The predicted academic performance of those who left school before the exam is much closer to the predicted academic performance of those who took the exam than in Table 4. This indicates that not including these four important variables in the model cannot be compensated by applying the usual econometric techniques, such as a Heckman selection model or a joint model including a selection and academic performance equation.

## V. Conclusion

In this paper we have provided empirical evidence on the effect that personal characteristics and family resources have on the academic performance of young persons and on their choices about dropping out of secondary school before taking national School Certificate Examinations.

The paper has addressed three questions of relevance to the analysis of academic performance. First, whether the model for Academic Performance of those taking the exam is representative of the complete student population including those who left school early. Second, whether the predicted Academic Performance of those leaving school early differs from the predicted academic performance of those taking the exam. Third, whether those variables that are often not observed in other datasets, but which are observed in the CHDS, are important for the results.

The extensive set of variables in and the longitudinal nature of the Christchurch Health and Development data set have allowed modelling and hypothesis testing of a number of relevant factors. In particular, the analysis of academic performance (as measured by an average national exam score at age 15) incorporated the effect of academic ability (as measured by IQ score at age eight), as well as school and peer effects, and household economic conditions over time. The study extends the literature by addressing the potential correlation between the unobserved terms in the academic performance and early school-leaving equations.

The analysis shows that those who left school early would have been expected to perform poorly on the national exam if they had taken the exam. The academic performance can be adequately predicted using the model estimated based on the students taking the exam. Unobserved characteristics do not appear to be relevant. The much poorer predicted performance of those leaving school compared to those staying to take the exam is nearly completely driven by observed characteristics.

However, it is noteworthy, that results change considerably when important characteristics such

---

<sup>13</sup> The log likelihood value for the model where the correlation is restricted to be equal to 0 is -842.16, versus a value of -842.08 for the full model including an estimated correlation term. We cannot reject the restricted model.

as, childhood IQ, early childhood family resources, association with deviant peers at age 15 and the proportion of the individual's 10<sup>th</sup> grade who go on to the 11<sup>th</sup> grade, are excluded from the model. These variables are often unavailable in other datasets. Although the error correlation coefficient remains insignificant, using this limited variable model allows much less differentiation in results. This leads to a predicted academic performance of early school leavers, which is much closer to the academic performance of those who stayed. This highlights that econometric specification is not a sufficient substitute for the presence of important personal, family and peer variables in the data on which academic performance models are based. The results presented in this paper show the importance of controlling for personal and family heterogeneity, as shown by the effect of the additional information available in the New Zealand CHDS.

We have shown that the inclusion of personal, family and peer information has the potential to influence the conclusions drawn from studies. For example, in this paper, the conclusion from the model with the limited number of variables could have been that a policy aimed at keeping students at school might be effective in raising the overall education level of the student population. The extensive model, including childhood and peer effects, however indicates that keeping students at school would not be enough. Even if those who are now leaving school would stay, they would be very likely to fail the national exam. The results from the model indicate that more needs to be done in terms of policies, keeping students at school would only be a first step.

**Table 1: Characteristics of Sample**

Characteristics	Mean (Standard Deviation)	
	Full Sample	Took National Exams
<b>Personal Characteristics</b>		
Female (%)	50.5%	52.0%
Maori Ethnicity (%)	7.4%	6.3%
Pacific Island Ethnicity (%)	2.8%	2.6%
IQ (tested at 8 years of age)	102.8 (14.88)	105.12 (13.70)
<b>Education</b>		
Average School Certificate Grade Point Average (age 15 and 10 <sup>th</sup> grade, where Fail=0, C=1, B=2, A=3)	1.061 (0.86)	1.21 (0.81)
Mother without Qualifications (<10 <sup>th</sup> grade)	49.8%	45.3%
Mother with a Tertiary Qualification	20.6%	23.4%
Father without Qualifications (< 10 <sup>th</sup> grade)	47.5%	42.8%
Father with a Tertiary Qualification	19.8%	22.2%
Total Dropout rate from school at Age 16	15.5%	---
Dropout rate from school before Exams	8.8%	---
<b>Family and Social Environment</b>		
Adolescent Average Income Decile: Ages 11-14 (10 is most affluent decile)	5.53 (2.54)	5.83 (2.50)
Early Childhood Average Income Decile: Ages 1-5	5.82 (2.40)	6.10 (2.33)
Own their Home (%)	88.6%	92.1%
Number of Siblings	1.48 (0.94)	1.49 (0.88)
Rural Location (%)	15.9%	16.0%
Percentage of Family Income from Benefits	13.9%	10.6%
Regional Unemployment Rate	10.6%	10.6%
Percentage of Respondent's class continue at Age 16	83.6% (16.2)	86.0 (11.6)
Average Class Size	28.8 (4.20)	28.9 (4.14)
Association with Deviant Peers age 15 (10 is the highest association)	2.30 (2.45)	2.00 (2.24)
<b>Sample Size:</b>	713	598

**Table 2: Academic Performance<sup>a</sup> (Tobit Specification)**

Explanatory Variables	With Selection Equation			No Selection Equation		
	Coefficie	t-ratio	P> t	Coefficie	z	P> z
<b>Average grade</b>						
Female	0.2361	4.28	0.000	0.2338	4.36	0.000
Maori	0.0933	0.73	0.465	0.0934	0.84	0.403
Pacific_Island	-0.0992	-0.47	0.636	-0.1051	-0.60	0.549
Mother_No_Qualifications	-0.1460	-2.20	0.028	-0.1425	-2.21	0.027
Mother_Higher_Qualifications	0.1205	1.66	0.098	0.1229	1.64	0.101
Father_No_Qualifications	-0.0384	-0.61	0.542	-0.0340	-0.54	0.588
Father_Higher_Qualifications	0.1313	1.70	0.089	0.1336	1.69	0.091
Number_Siblings	0.0007	0.02	0.982	0.0004	0.01	0.990
Own_Home	-0.1159	-0.98	0.326	-0.1226	-1.15	0.252
Rural	0.0329	0.48	0.634	0.0320	0.43	0.667
Welfare_Benefit_Proportion	0.1767	1.48	0.140	0.1856	1.65	0.100
Inc_Decile (ages 11-14)	0.0598	3.92	0.000	0.0603	4.02	0.000
Early_Inc_Decile(ages1-5)	0.0394	2.73	0.006	0.0387	2.72	0.007
IQ8	0.0298	12.04	0.000	0.0295	14.01	0.000
Proportion of Class_Continue	0.6201	2.16	0.030	0.5796	2.36	0.018
Peer_Deviant	-0.0665	-3.99	0.000	-0.0639	-5.22	0.000
Constant	-2.9699	-6.95	0.000	-2.8998	-9.23	0.000
<b>Stayed at school</b>						
Female	0.1848	0.87	0.384			
Maori	0.1149	0.40	0.687			
Pacific_Island	0.2373	0.46	0.647			
Mother_No_Qualifications	-0.4133	-1.92	0.055			
Father_No_Qualifications	-0.3909	-1.77	0.077			
Number_Siblings	0.1106	1.04	0.298			
Own_Home	0.4186	1.77	0.077			
Rural	0.1670	0.40	0.692			
Welfare_Benefit_Proportion	-0.1461	-0.46	0.649			
Local_Unemployment	0.1873	0.58	0.565			
Inc_Decile (ages 11-14)	-0.0012	-0.02	0.986			
Early_Inc_Decile(ages1-5)	0.0678	1.61	0.107			
IQ8	0.0400	4.29	0.000			
Proportion of Class_Continue	2.7255	4.59	0.000			
Peer_Deviant	-0.2110	-6.02	0.000			
Ave_Class Size	0.0272	1.05	0.295			
Birth Month is April/May	0.3828	1.67	0.095			
Constant	-7.3079	-1.95	0.051			
Variance of the error term	0.6346	27.68	0.000	0.6344	31.88	0.000
Correlation (ρ)	0.1184	0.21	0.837			
Number of observations = 661 (598 with observed academic performance)			LR chi <sup>2</sup> (16) =332.1	Number of observations = 598		
Log likelihood = -684.3286			Prob > chi <sup>2</sup> = 0.0000	Log likelihood = -580.3863		

Note a: Academic Performance: Average Grade in National Examinations in Grade 10

Selection Equation: Stayed at school: 1=stayed at school until at least the exam; 0=left school before the exam

**Table 3: Academic Performance<sup>a</sup> (OLS Specification)**

Explanatory Variables	With Selection Equation			No Selection Equation		
	Coefficient	t-ratio	P> t	Coefficient	t	P> t
<b>Average grade</b>						
Female	0.2139	4.37	0.000	0.2082	4.24	0.000
Maori	0.0503	0.50	0.618	0.0501	0.49	0.624
Pacific_Island	-0.1061	-0.67	0.501	-0.1222	-0.77	0.442
Mother_No_Qualifications	-0.1369	-2.31	0.021	-0.1274	-2.16	0.031
Mother_Higher_Qualifications	0.1175	1.71	0.087	0.1242	1.80	0.073
Father_No_Qualifications	-0.0403	-0.69	0.488	-0.0286	-0.50	0.619
Father_Higher_Qualifications	0.1233	1.70	0.088	0.1304	1.79	0.074
Number_Siblings	0.0116	0.42	0.673	0.0110	0.40	0.691
Own_Home	-0.0876	-0.90	0.370	-0.1053	-1.08	0.280
Rural	0.0385	0.57	0.566	0.0361	0.53	0.594
Welfare_Benefit_Proportion	0.1957	1.90	0.058	0.2200	2.17	0.031
Inc_Decile (ages 11-14)	0.0554	4.04	0.000	0.0564	4.09	0.000
Early_Inc_Decile(ages1-5)	0.0352	2.71	0.007	0.0334	2.57	0.011
IQ8	0.0277	13.49	0.000	0.0269	14.05	0.000
Proportion of Class_Continue	0.6747	2.74	0.006	0.5695	2.54	0.011
Peer_Deviant	-0.0623	-4.82	0.000	-0.0556	-5.02	0.000
Constant	-2.7598	-8.08	0.000	-2.5738	-9.00	0.000
<b>Stayed at school</b>						
Female	0.1772	0.72	0.472			
Maori	0.1178	0.38	0.701			
Pacific_Island	0.2808	0.58	0.562			
Mother_No_Qualifications	-0.4061	-1.73	0.084			
Father_No_Qualifications	-0.3924	-1.73	0.083			
Number_Siblings	0.1062	1.07	0.285			
Own_Home	0.4266	1.60	0.109			
Rural	0.1621	0.38	0.701			
Welfare_Benefit_Proportion	-0.1486	-0.50	0.616			
Local_Unemployment	0.1851	0.45	0.653			
Inc_Decile (ages 11-14)	0.0032	0.05	0.960			
Early_Inc_Decile(ages1-5)	0.0673	1.31	0.191			
IQ8	0.0399	4.61	0.000			
Proportion of Class_Continue	2.6775	5.04	0.000			
Peer_Deviant	-0.2088	-5.52	0.000			
Ave_Class Size	0.0281	1.14	0.256			
Birth Month is April/May	0.3962	1.88	0.061			
Constant	-7.2980	-1.60	0.109			
Variance of the error term	0.5829					
Correlation ( $\rho$ )	0.3289					
Mills ratio	0.1917	0.98	0.329			
Number of observations = 661	Wald $\chi^2(30) = 490.79$			Number of observations = 598		
	Prob > $\chi^2 = 0.0000$			Adj R-squared = 0.4728		

Note a: Academic Performance: Average Grade in National Examinations in Grade 10

Selection Equation: Stayed at school: 1=stayed at school until at least the exam; 0=left school before the exam

**Table 4: Predicted Average Grade for Full Sample, Based on Different Specifications With and Without Controlling for Sample Selection**

<b>Variable</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>Min</b>	<b>Max</b>
Group 1: Did Not drop-out before, and took exams (n=598):				
Average_Grade* using Tobit	1.1751	0.6208	0.0000	3.0000
Average_Grade using Tobit	1.1815	0.6137	0.0000	3.0000
Average_Grade* using OLS	1.1921	0.5869	-0.2276	2.8931
Average_Grade using OLS	1.2103	0.5651	-0.1490	2.8615
Group 2: Dropped out before exams, and Did Not take exams (n=60)				
Average_Grade* using Tobit	0.2468	0.3716	0.0000	1.3481
Average_Grade using Tobit	0.2635	0.3782	0.0000	1.3986
Average_Grade* using OLS	0.1393	0.5394	-1.0623	1.3210
Average_Grade using OLS	0.2364	0.5133	-0.8368	1.4223

Note: Average\_Grade\* controls for sample selection, while Average\_Grade does not.

**Table 5: Academic Performance<sup>a</sup> (Tobit Specification with Fewer Explanatory Variables)**

Explanatory Variables	With Selection Equation			No Selection Equation		
	Coefficient	t-ratio	P> t	Coefficient	z	P> z
<b>Average grade</b>						
Female	0.1720	2.68	0.007	0.1757	2.76	0.006
Maori	0.0759	0.52	0.605	0.0725	0.55	0.584
Pacific_Island	-0.3510	-1.77	0.076	-0.3490	-1.66	0.096
Mother_No_Qualifications	-0.1912	-2.43	0.015	-0.1976	-2.57	0.010
Mother_Higher_Qualifications	0.2578	3.10	0.002	0.2567	2.89	0.004
Father_No_Qualifications	-0.1175	-1.55	0.121	-0.1250	-1.68	0.093
Father_Higher_Qualifications	0.3003	3.57	0.000	0.2988	3.25	0.001
Number_Siblings	-0.0139	-0.40	0.687	-0.0146	-0.40	0.687
Own_Home	-0.0191	-0.14	0.891	0.0009	0.01	0.994
Rural	0.0635	0.72	0.471	0.0630	0.72	0.474
Welfare_Benefit_Proportion	0.1516	0.99	0.323	0.1387	1.04	0.298
Inc_Decile (ages 11-14)	0.1023	6.32	0.000	0.1035	6.31	0.000
Early_Inc_Decile(ages1-5)	--					
IQ8	--					
Proportion of Class_Continue	--					
Peer_Deviant	--					
Constant	0.5294	2.69	0.007	0.5004	2.79	0.005
<b>Stayed at school</b>						
Female	0.1813	0.94	0.345			
Maori	-0.0735	-0.26	0.791			
Pacific_Island	0.0955	0.21	0.831			
Mother_No_Qualifications	-0.4475	-2.56	0.011			
Father_No_Qualifications	-0.5195	-2.97	0.003			
Number_Siblings	-0.0533	-0.62	0.535			
Own_Home	0.7363	3.66	0.000			
Rural	0.2033	0.58	0.565			
Welfare_Benefit_Proportion	-0.2684	-1.08	0.279			
Local_Unemployment	0.2656	0.86	0.392			
Inc_Decile (ages 11-14)	0.1168	2.65	0.008			
Early_Inc_Decile(ages1-5)	--					
IQ8	--					
Proportion of Class_Continue	--					
Peer_Deviant	--					
Ave_Class Size	0.0104	0.50	0.614			
Birth Month is April/May	0.3014	1.75	0.080			
Constant	-2.4263	-0.72	0.473			
Variance of the error term	0.7606	29.68	0.000	0.7599	31.72	0.000
Correlation (ρ)	-0.0974	-0.85	0.393			
Number of observations = 661 (598 with observed academic performance)		LR =200.0 chi <sup>2</sup> (12)		Number of observations = 598		
Log likelihood = -842.0833		Prob > chi <sup>2</sup> = 0.0000		Log likelihood = -683.6447		

Note a: Academic Performance: Average Grade in National Examinations in Grade 10

Selection Equation: Stayed at school: 1=stayed at school until at least the exam; 0=left school before the exam

**Table 6: Academic Performance<sup>a</sup> (OLS Specification with Fewer Explanatory Variables)**

Explanatory Variables	With Selection Equation			No Selection Equation		
	Coefficient	t-ratio	P> t	Coefficient	t	P> t
<b>Average grade</b>						
Female	0.1385	2.23	0.026	0.1545	2.67	0.008
Maori	0.0481	0.40	0.693	0.0340	0.28	0.778
Pacific_Island	-0.3261	-1.74	0.082	-0.3166	-1.69	0.092
Mother_No_Qualifications	-0.1514	-1.90	0.058	-0.1793	-2.57	0.010
Mother_Higher_Qualifications	0.2473	3.04	0.002	0.2427	2.99	0.003
Father_No_Qualifications	-0.0762	-0.93	0.352	-0.1094	-1.62	0.106
Father_Higher_Qualifications	0.2865	3.40	0.001	0.2794	3.33	0.001
Number_Siblings	-0.0003	-0.01	0.992	-0.0034	-0.10	0.917
Own_Home	-0.0837	-0.50	0.617	0.0052	0.05	0.964
Rural	0.0690	0.86	0.389	0.0672	0.84	0.400
Welfare_Benefit_Proportion	0.2225	1.56	0.119	0.1655	1.38	0.168
Inc_Decile (ages 11-14)	0.0898	5.36	0.000	0.0952	6.38	0.000
Early_Inc_Decile(ages1-5)	--					
IQ8	--					
Proportion of Class_Continue	--					
Peer_Deviant	--					
Constant	0.6893	2.89	0.004	0.5614	3.45	0.001
<b>Stayed at school</b>						
Female	0.1805	0.92	0.358			
Maori	-0.0766	-0.30	0.764			
Pacific_Island	0.0833	0.21	0.837			
Mother_No_Qualifications	-0.4441	-2.47	0.013			
Father_No_Qualifications	-0.5182	-2.96	0.003			
Number_Siblings	-0.0511	-0.64	0.523			
Own_Home	0.7330	3.52	0.000			
Rural	0.2112	0.58	0.561			
Welfare_Benefit_Proportion	-0.2656	-1.12	0.262			
Local_Unemployment	0.2794	0.86	0.390			
Inc_Decile (ages 11-14)	0.1171	2.59	0.010			
Early_Inc_Decile(ages1-5)	--					
IQ8	--					
Proportion of Class_Continue	--					
Peer_Deviant	--					
Ave_Class Size	0.0091	0.45	0.650			
Birth Month is April/May	0.2910	1.73	0.083			
Constant	-2.5366	-0.72	0.469			
Variance of the error term	0.7043					
Correlation ( $\rho$ )	-0.4615					
Mills ratio	-0.3250	-0.73	0.464			
Number of observations = 661	Wald chi <sup>2</sup> (22) = 199.06			Number of observations = 598		
	Prob > chi <sup>2</sup> = 0.0000			Adj R-squared = 0.2583		

Note a: Academic Performance: Average Grade in National Examinations in Grade 10

Selection Equation: Stayed at school: 1=stayed at school until at least the exam; 0=left school before the exam

**Table 7: Predicted Average Grade for Full Sample, based on Different Specifications with and without controlling for Sample Selection (as specified in Tables 5 and 6)**

<b>Variable</b>	<b>Mean</b>	<b>Std. Dev.</b>	<b>Min</b>	<b>Max</b>
Group 1: Did Not drop-out before, and took exams (n=598):				
Average_Grade* using Tobit	1.1893	0.4541	0.0767	2.3126
Average_Grade using Tobit	1.1786	0.4623	0.0293	2.3159
Average_Grade* using OLS	1.2572	0.3892	0.3658	2.2445
Average_Grade using OLS	1.2103	0.4233	0.1726	2.2595
Group 2: Dropped out before exams, and Did Not take exams (n=60)				
Average_Grade* using Tobit	0.8588	0.3285	0.5274	1.1844
Average_Grade using Tobit	0.8406	0.3470	0.4883	1.1822
Average_Grade* using OLS	0.9706	0.2518	0.7311	1.2332
Average_Grade using OLS	0.8904	0.3323	0.5585	1.2231

Note: Average\_Grade\* controls for sample selection, while Average\_Grade does not.

## Bibliography

- Becker, GS (1993), *Human Capital*, third edition, Chicago University Press.
- Blau, DM (1999), "The Effect of Income on Child Development", *The Review of Economics and Statistics*, 81 (2), 261-276.
- Borjas, GJ (1995), "Ethnicity, Neighbourhoods, and Human-Capital Externalities", *American Economic Review*, 85(3), 365-390.
- Borland, J and Wilkins, R (1996), "Earnings Inequality in Australia", *Economic Record*, 72(216), 7-23.
- Case, AC and Katz, LF (1991), "The Company You Keep: The Effects of Family and Neighbourhood on Disadvantaged Youths", NBER Working Paper 3705.
- Duncan, GJ, Yeung, WJ, Brooks-Gunn, J and Smith, J (1998), "How Much Does Child Poverty Affect the Life Chances of Children?", *American Sociological Review*, 63, 406-23.
- Ermisch, J and Francesconi, M (2001), "Family Matters: Impact of Family Background on Educational Attainment", *Economica*, May, 137-156.
- Feinstein, L and Symons, J (1999), "Attainment in Secondary School", *Oxford Economic Papers*, 51, 300-321.
- Fergusson, DM, Horwood, LJ and Lloyd, M (1991), "Confirmatory Factor Models of Attention Deficit and Conduct Disorder", *Journal of Child Psychology and Psychiatry*, 32(2), 257-274.
- Fergusson, DM, Horwood, LJ, Shannon, FT and Lawton, JM (1989), "The Christchurch Child Development Study: A Review of Epidemiological Findings", *Paediatric and Perinatal Epidemiology*, 3, 302-325.
- Fergusson, DM and Lynskey, MT (1993), "Maternal Age and Cognitive and Behavioural Outcomes in Middle Childhood", *Paediatric and Perinatal Epidemiology*, 7, 77-91.
- Gregg, P. and Machin, S. (1998), "Child Development and Success or Failure in the Youth Labour Market", Centre for Economic Performance Discussion Paper: no 397, London School of Economics.
- Haveman, R and Wolfe, B (1995), "The Determinants of Children's Attainment: A Review of Methods and Findings", *Journal of Economic Literature*, 33, 1829-78.
- Maani, SA (1996), *A Research Agenda, Methodologies, Models and Data Requirements for Estimating Participation in Post-compulsory Education in New Zealand*, Report prepared for the New Zealand Treasury, Wellington.
- Maani, SA (1997), *Investing in Minds: The Economics of Higher Education in New Zealand*, Institute of Policy Studies, Wellington.
- Maani, SA (2006), "Parental Income and the Choice of participation in University, Polytechnic, or Employment at Age Eighteen: A Longitudinal Study", in *Research in Economic Inequality*,

*Volume 13*, eds. J. Creedy & G. Kalb, Elsevier, 217-248.

Maani, S.A and Kalb, G. (forthcoming, 2007), "Academic performance, Childhood Economic Resources, and the Choice to Leave School at Age Sixteen", *Economics of Education Review*, 26(2).

Maloney, T. (1999), "An Analysis of Representativeness and Attrition in the CHDS", New Zealand Treasury Working Paper , September 1999.

Miller, P and Volker, P (1989), "Socio-economic Influences on Educational Attainment: Evidence and Implications for the Tertiary Education Finance Debate", *Australian Journal of Statistics*, 31A, 47-70.

Montgomery, J (1991), "Social Networks and Labour-Market Outcomes: Toward an Economic Analysis", *American Economic Review*, 81(5), 1408-1418.

Prior, H and Beggs, JJ (1989), "Influence of Family Background on the Educational and Labour-Force Outcomes of Year 12 School-Leavers", *Australian Journal of Statistics*, 31A, 99-124.

Rice, PG (1987), "The Demand for Post-Compulsory Education in the UK and the Effects of Educational Maintenance Allowances", *Economica*, 54 465-475.

Schultz, TW (1961), "Investment in Human Capital", *American Economic Review*, 51, 1-17.

Willis, RJ and Rosen, S (1979), "Education and Self-Selection", *Journal of Political Economy*, 87, S7-S36.

## APPENDIX A

**Table A1: Definition of Variables**

Stayed at School	Binary dependent variable: Stayed at school: 1=stayed at school until at least the exam; 0=left school before the examination.
Average School Certificate Grade at Age 15 (Ave_Grade)	The average value of all School Certificate (10 <sup>th</sup> grade, age 15) Examination Marks over all subjects taken with weightings of 3 for an A, 2 for a B, 1 for a C and 0 for a fail (D).
Proportion of Students Continuing (Proportion of Class_Continue)	Proportion of an individual's 10 <sup>th</sup> grade (Fifth Form) class within the data set continuing onto the 11 <sup>th</sup> grade. The relevant individual is excluded from the calculation.
Female	Binary=0 for a male; 1 for a female.
Maori	Binary=1 if Maori.
Pacific Islander (Pacific_Island)	Binary=1 if a Pacific Islander.
Mother without Qualifications (Mother_No_Qualifications)	Binary=1 if child's mother does not have formal educational qualifications (10 <sup>th</sup> grade School Certificate or higher).
Mother with Tertiary Qualifications (Mother_Higher_Qualifications)	Binary=1 if a child's mother has a university or other tertiary qualification.
Father without Qualifications (Father_No_Qualifications)	Binary=1 if a child's father does not have formal educational qualifications (10 <sup>th</sup> grade or higher).
Father with Tertiary Qualifications (Father_Higher_Qualifications)	Binary=1 if child's father has a university or other tertiary qualification.
Number of Siblings (Number_Siblings)	Number of siblings in the home at 15 years.

**Table A1: continued**

Parents Own their Own Home (Own_Home)	Binary=1 if parents own their own home and the child is living at home at 15 years of age.
Rural Lifestyle (Rural)	Binary=1 if a child was not living in a main urban centre at 15 years of age.
Proportion of Family Income from Benefits (Welfare_Benefit_Proportion)	The proportion (between 0 and 1) of the family's income derived from social welfare benefits. This variable is expected to reflect relative disadvantage in terms of parental assets, relative income, and other disadvantage in terms of information or social networks. <sup>14</sup>
Registered Unemployment (Local_Unemployment)	Regional unemployment rate by gender in which each individual was living at 15 years of age. <sup>15</sup> There were 8 regions and the corresponding levels of unemployment ranged between 5.9 and 12.1 percent.
Average Income Decile (Inc_Decile)	Average income decile of the family when the child was between ages 11 and 14: 1 is consistently poor; 10 is consistently affluent.
Early Childhood Average Income Decile (Early_Inc_Decile)	Average income decile of the family when the child was between ages 1 and 5: 1 is consistently poor; 10 is consistently affluent.
Total Intelligence Quotient (IQ8)	The child's measured total IQ score at 8 years of age (revised Wechsler Intelligence Scale for Children).
Class Size (Class_Size)	Average class size in secondary school
Affiliation with Deviant Peers (Peer_Deviant)	Affiliation with deviant peers at age 15 based upon self-reported use of tobacco, alcohol, illicit drugs, other illegal behaviour, etc. by friends: 0-10, with 10 being the most deviant affiliations.

---

<sup>14</sup> The use of various potential benefits received by the young persons themselves did not prove useful since all respondents were potentially eligible for the unemployment benefit, or a training benefit. For example, the receipt of the unemployment benefit by 6.5% of the total sample was itself a result of unemployment 'choices' and therefore not a relevant independent predictor. The same is true of the Training Benefit, which was received by 5.2% of the sample who had taken part in training. The Independent Youth benefit could in turn be received independently of education or employment choices and was potentially relevant. This benefit was received by only 1 percent of the sample and it was not statistically significant.

<sup>15</sup> Source of this information is the 1991 Census of Population and Dwellings: Regional Summary.