THE UNIVERSITY OF MELBOURNE

Melbourne Institute Working Paper Series

Working Paper No.5/07

The Importance of Observing Early School Leaving and Usually Unobserved Background and Peer Characteristics in Analysing Academic Performance

Guyonne Kalb and Sholeh A. Maani

MELBOURNE INSTITUTE
of Applied Economic and Social Research

The Importance of Observing Early School Leaving and Usually Unobserved Background and Peer Characteristics in Analysing Academic Performance*

Guyonne Kalb[†] and Sholeh A. Maani[‡]

[†] Melbourne Institute of Applied Economic and Social Research, The University of Melbourne

[‡] Department of Economics, The University of Auckland

Melbourne Institute Working Paper No. 5/07

ISSN 1328-4991 (Print)

ISSN 1447-5863 (Online)

ISBN 978-0-7340-3235-5

February 2007

* This study was made possible through a research grant from the New Zealand Treasury. Special thanks go to David Fergusson and John Horwood for making the data set available and for insightful information on the data set; Francine Blau, Lawrence Kahn, Stephen Jenkins, Anna Vignoles, Gita Kingdon, Bob Haveman, Barbara Wolfe, John Karl Scholz and Geoffrey Wallace, and the participants of the 2006 Australian Labour Econometrics Workshop for insightful comments; George Barker and Ron Crawford for helpful comments on an earlier part of the study; and Adam Warner for initial research assistance. The authors alone are responsible for the views expressed.

Melbourne Institute of Applied Economic and Social Research

The University of Melbourne

Victoria 3010 Australia

***Telephone* (03) 8344 2100**

***Fax* (03) 8344 2111**

***Email* melb-inst@unimelb.edu.au**

***WWW Address* http://www.melbourneinstitute.com**

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 by teenagers to drop out of school before exams 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 endogeneity and error correlation of potential performance in national examinations and school-leaving choices prior to exams. Birth month provides an instrument 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. The analyses incorporate the effect of academic ability (childhood IQ), 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. The predicted difference between those who drop out or 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, childhood family resources and teenage peer effects), we find that the unobserved factors in academic performance and early school leaving are correlated. It is found that beyond childhood IQ and family resources, teenage peer and school factors have additional and significant associations with grade outcomes. This has important policy implications.

1. 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, it examines 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. This is of additional interest, in allowing us to observe both early childhood ability and family resources, and teenage peer performance characteristics.

Demand for post-compulsory education as a personal choice has been analysed 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 of interest internationally. The question of the effect of parental resources on academic performance of children and adolescents and their later schooling outcomes has 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 substantial proportion of students starts school later and would turn 16 before the exams. Therefore, they could leave school legally before the exams take place.

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[1], 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 2. A discussion of the data set and the characteristics of the sample follow

[1] 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).

in Section 3. The estimated models and their results are presented in Section 4, followed by concluding remarks in Section 5.

2. 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).[2]

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 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

[2] A detailed review of the literature on participation in post-compulsory and higher education is provided in

and emphasised the effect of financial constraints on school leaving choices of males and females.[3] Neither study had observable variables on academic ability such as IQ or academic test scores. Micklewright's (1989) study used childhood maths and reading scores, and parental current income in examining schooling choices at age sixteen, in the UK.

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

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

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:

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

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

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

with β, γ and σ^2 constant across the population (e.g. Willis and Rosen, 1979; Rice, 1987).

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

Maani (1997).

[3] 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).

$A_i = f(S_i, X_i, v_i)$ (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.[4] The recent literature on educational attainment has emphasised 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 (see for example, Ermisch and Francesconi, 2001).[5] 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 4, 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.

---

4 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.

5 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. Carniero and Heckman (2002) consider the effect of credit constraints of youth on their College enrolment.

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. For example, we are able to control for both early childhood IQ and family income throughout early childhood and teenage years, and teenage peers' school performance and characteristics. 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.

3. 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.[6] The characteristics of this sample are summarised in Table 1 below. 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.).

6 The original cohort of individuals in the survey consisted of 1265 individuals. The sample used in this study contains 661 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.

Table 1: Characteristics of the Sample

Characteristics	Mean (Standard Deviation) Full Sample	Took National Exams
Personal Characteristics		
Female (%)	51.4%	52.0%
Maori Ethnicity (%)	7.1%	6.3%
Pacific Island Ethnicity (%)	2.7%	2.6%
IQ (tested at 8 years of age)	103.6 (14.47)	105.12 (13.70)
Education		
Average School Certificate Grade Point Average (age 15 and 10th grade, where Fail=0, C=1, B=2, A=3)	1.09 (0.85)	1.21 (0.81)
Mother without Qualifications (<10th grade)	48.6%	45.3%
Mother with a Tertiary Qualification	21.1%	23.4%
Father without Qualifications (< 10th grade)	46.1%	42.8%
Father with a Tertiary Qualification	20.3%	22.2%
Total Dropout rate from school at Age 16	15.5%	---
Dropout rate from school before Exams	9.5%	---
Family and Social Environment		
Adolescent Average Income Decile: Ages 11-14 (10 is most affluent decile)	5.60 (2.55)	5.83 (2.50)
Early Childhood Average Income Decile: Ages 1-5	5.89 (2.40)	6.10 (2.33)
Own their Home (%)	89.1%	92.1%
Number of Siblings	1.48 (0.92)	1.49 (0.88)
Rural Location (%)	16.3%	16.0%
Percentage of Family Income from Benefits	13.6%	10.6%
Regional Unemployment Rate	10.6%	10.6%
Percentage of Respondent's class continue at Age 16	83.7% (16.3)	86.0 (11.6)
Average Class Size	28.8 (4.12)	28.9 (4.14)
Association with Deviant Peers age 15 (10 is the highest association)	2.28 (2.45)	2.00 (2.24)
Sample Size:	661	598

As column 1 of Table 1 on the full sample shows, about half of the sample (51.4%) 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 103.6, and the average school certificate mark of 1.09 or a C. Home ownership by parents was 89.1%, and the average proportion of family income from benefits was 13.6%. In the sample, 7.1% were Maori and 2.7% were Pacific Islanders. On parental education, 48.6% of the mothers and 46.1% of the fathers of the respondents had no school qualifications (less than the year 10 School Certificate), and 21.1% of mothers and 20.3% of fathers had tertiary qualifications.

Column 2 in Table 1 presents the mean characteristics for those who had not dropped out before exams. 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 early school leavers. These differences include a higher average IQ at age 8, they belonged to a higher family income decile, and they went to a school with a higher proportion of the class continuing to the Sixth Form. These characteristics are consistent with the hypothesis that individuals sort themselves into different choices based on their academic ability, 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.

4. Models and Results

In this section, the occurrence 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. We use 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.

4.1 School Leaving

The school-leaving 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.[7]

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.

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 importance 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.[8]

---

7 See Table A1 in Appendix A for a full list of variables and their descriptions.

8 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.

4.2 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. For each subject, 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.[9] 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 or 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:

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

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

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

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

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

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 within a five-month interval of April through August. Those whose birth dates fall in semester 1 (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 whose birth dates fall in semester 2 (June to August) start school in the second semester of the year, when they turn five, and

9 This is also compatible with the official GPA (Grade Point Average) score assignment in New Zealand.

generally 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 in the earlier months of the year turn 16 only after the exams. Therefore, those in the CHDS who are a few months older due to being born in the first semester are more likely to be at school at the time of examinations. Birth month is expected to affect school leaving, but not academic performance. This is empirically verified in our sample, and it is used to identify the models in this paper.

In our sample, 38.1% of the births fell in the last part of the first semester (April-May), and the rest of the births fell in the early part of the second semester (June-August). Table A2 in the Appendix provides a summary of mean characteristics by semester of birth in our sample. These mean characteristics show that the two groups were not statistically different (e.g. in childhood IQ, family income, age 15 exam grades, and school characteristics), as confirmed by t-tests across the two sub-samples. In addition, a higher percentage of those who were born in the second semester, and could legally leave before exams, had left school compared to those who were born in the first semester. That is, 3.6% of those born in April (which lies clearly in Semester 1) left school before exams, compared to 14.6% for those born in August (which lies clearly in Semester 2)—a difference that is statistically significant. Likewise, 7.5% of those who were born in the combined months of Semester 1, compared to 11.0% for birth months falling in Semester 2 on average left school before exams. Although this combined raw difference is not statistically significant (possibly due to the small sample size, or some potential flexibility for those born between semesters), this difference in the average proportion leaving school before exams is larger and closer to significance than any of the other differences in Table A2. This characteristic of our data provides useful counterfactual observations of students who are comparable otherwise, except for the ability to leave school legally before the exams due to having turned sixteen. Interestingly, the fact that all births fall within five months, decreases potential age impacts on learning noted with quarter of birth (e.g. in Angrist and Krueger, 1992).

We confirm the validity of the birth month variable as an exclusion restriction for joint estimations. We find no evidence of a direct effect by the birth month binary variable in the academic performance model. This is consistent across single-equation estimates, and joint estimations (the latter is identified through functional form of the Tobit and the Probit), and all corresponding tests of restrictions.[10]

4.2.1 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. 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.

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.

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 early childhood and adolescent years. Some respondents experienced improvements in their household's relative income position, while others experienced deteriorations. Including both teenage and early childhood income decile variables, we find that they are both 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 explains an additional effect of 0.039 of a grade per decile. 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

---

10 When included in our single equation or joint academic performance models, the Birth Month_Semester 1 variable is highly insignificant. Likewise, the log likelihood value for the jointly estimated model where the birth month coefficient is set to zero in the academic performance equation is -684.32, versus a value for the full model, including the variable, of -684.24. We cannot reject the restricted model.

deciles difference (the difference between a C or a D average grade, for example).

Table 2: Academic Performance[a] (Tobit Specification)

Explanatory Variables	With Selection Equation			No Selection Equation		
Average grade	Coefficient	z-value	P> |z|	Coefficient	z-value	P>|z|
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
Stayed at school						
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_Semester 1	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			
No. of obs. = 661 (598 with obs. average grade)		LR $chi^2(16)$ =332.1		Number of observations =		598
Log likelihood = -684.3286		Prob > chi^2 = 0.0000		Log likelihood = -580.3863		

Note a: Academic Performance: Average Grade in National Examinations in Grade 10
Selection Equation: 1=stayed at school until at least the exam; 0=left school before the exam

These results are consistent with the effect of poverty (Duncan et al., 1998), 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.

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.[11]

[11] There are two further variables in our model, which are included in the probit selection equation only, as they are consistently insignificant in the academic performance models. Local_ Unemployment rate is expected to affect school leaving directly due to job opportunities, but not necessarily academic performance. Average_Class Size during secondary school years is also insignificant in all our academic performance models. However, the variables did not have statistically significant effects in the school-leaving equation either.

The correlation coefficient ρ 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 (also in Table 2), 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. [12]

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.[13] 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.

---

12 This result is consistent with Card and Rothstein's (2005) conclusion that in general the absence of schoolmate characteristics would lead to overestimation of negative ethnicity effects on academic achievement.

13 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:

$$\Phi^{-1}(P_{di}) = Z_i{}'\gamma_Z \qquad (9)$$

The estimated parameters γ_Z 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:

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

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

ϕ is the standard normal probability density function and Φ is the standard normal cumulative distribution function.

Table 3: Academic Performance[a] (OLS Specification)

Explanatory Variables	With Selection Equation			No Selection Equation		
Average grade	Coefficient	z-value	P>|z|	Coefficient	t-ratio	P>|t|
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
Stayed at school						
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_Semester 1	0.3962	1.88	0.061			
Constant	-7.2980	-1.60	0.109			
Variance of the error term	0.5829					
Correlation (ρ)	0.3289					
Mills ratio	0.1917	0.98	0.329			
Number of observations = 661	Wald $chi^2(30)$ = 490.79			Number of observations = 598		
(598 with obs. average grade)	Prob > chi^2 = 0.0000			Adjusted 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

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.[14] 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.

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

Variable	Mean	Std. Dev.	Min	Max
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.

The first row for both categories 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 take the exam. Controlling for sample selection makes the results slightly

[14] Using the estimated parameters from the dropout equation (γz), an estimate of λ 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 $\lambda_i = \dfrac{\varphi(Z_i{}'\gamma_Z)}{\Phi(Z_i{}'\gamma_Z)}$.

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.

4.2.2 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 and on the conclusion that would be drawn from the model's results. For this reason, we re-estimate the model in Table 2, 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^{th} grade who go on to the 11^{th} grade. The same sample is used for this analysis and for the analysis in 4.2.1, 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 Table 5. It shows a clear difference in results, with the error correlation now being more significant and of a negative sign. Other differences are that the effect of parental education and being from Pacific Islander descent are now more significant, as is the effect of teenage 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 correlation term is still not significant at conventional levels.[15]

---

[15] Similar results, not presented here, were obtained in the OLS specification.

Table 5: Academic Performance[a] (Tobit Specification with Fewer Explanatory Variables)

Explanatory Variables	With Selection Equation			No Selection Equation		
Average grade	Coefficient	z-value	P>|z|	Coefficient	z-value	P>|z|
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
Stayed at school						
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_Semester 1	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)
Log likelihood = -842.0833

LR $chi^2(12)$ =200.0
Prob > chi^2 = 0.0000

Number of observations = 598
Log likelihood = -683.6447

Note a: Academic Performance: Average Grade in National Examinations in Grade 10
Selection Equation: 1=stayed at school until at least the exam; 0=left school before the exam

Table 6 presents the predicted average scores based on the models in Table 5. The results are clearly different from those presented in Table 4. As before, the differences between the estimates that account for potential self-selection into taking the examination and the estimates 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 still not be rejected, it is clear much less has been captured by the observed heterogeneity than in the full model.[16] 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.

Table 6: Predicted Average Grade for Full Sample, Based on Different Specifications with and without Controlling for Sample Selection (as specified in Table 5)

Variable	Mean	Std. Dev.	Min	Max
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
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

Note: Average_Grade* controls for sample selection, while Average_Grade does not.

We further examine the separate effects of the four excluded variables by estimating alternative specifications. In particular, it is of interest to examine whether later teenage peer and school factors are also important on their own, and separate from the effects of childhood IQ and early childhood family income (see for example, Heckman and Krueger, 2003). In these additional specifications of the model, we first excluded only IQ, followed by excluding both IQ and early childhood family income decile. Table 4 shows that our full model predicts

16 The log likelihood value for the restricted model where the correlation is set 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.

a clear fail or a predicted exam score of 0.246 or a D for those who had left school before examinations. When only IQ is excluded, the joint Tobit model provides less differentiation in predictive power and predicts a score of 0.366. When both IQ and early childhood income are excluded, the model's predicted score increases marginally to 0.385. However, when (as in Tables 5 and 6) our two variables on teenage 'peer deviant behaviour', and the 'proportion of the rest of the class continuing to further education' are also excluded, the model's predictive power diminishes significantly further as it now predicts a grade of 0.859 or a near pass for those who did not take the exam. The exclusion of the latter two variables have similarly sized effects on the predicted academic performance of dropouts, with the 'proportion of the rest of the class continuing to further education' being slightly bigger. This clearly shows the importance of both the early childhood and later factors in predicting academic performance.

The above analyses show that despite controlling for important childhood factors and background characteristics in our model, teenage peer and school characteristics have significant additional associations with academic performance. This is of interest since these specific teenage measures are often unobserved in data sets underlying debates on the relative importance of early childhood versus later environmental and resource characteristics in determining educational outcomes (see for example, Heckman and Krueger, 2003). These results indicate the usefulness of a potentially wider range of policy options, addressing teenage school and peer effects on academic performance.

The inclusion of personal, family and teenage 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.

In addition, our results have implications for "No Child Left Behind" policies in the US through the importance of both early childhood and later environmental resources, and policy implications in the UK in identifying school and peer characteristics that contribute to academic attainment. These latter teenage variables that we have shown to be important may be influenced through policy on learning and behavioural aspects of teenagers' school and peer environments. For example, Rothstein (2004) considers the SAT score in the US in the prediction of college performance, as opposed to high school GPA as a predictor. Including information on the student's gender and ethnicity, and the (fraction) ethnicity of the high school attended, he finds that the SAT's role in prediction models may be quite sensitive to the inclusion of background variables. This is particularly true with high school peer ethnicity characteristics as predictors. Our results indicate that school ethnicity in Rothstein's study may have served as a proxy for factors which can be influenced through policy, such as school-peer achievement or peer behaviour, as opposed to ethnicity itself. Identifying these underlying factors is important to enable appropriate policy design.

5. 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 extensive set of variables 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 paper has addressed three questions of relevance to the analysis of academic performance. First, it examines 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, it compares whether the predicted Academic Performance of those leaving school early differs from the predicted academic performance of those taking the exam. Third, it checks whether those variables that are often not observed in other datasets, but which are observed in the CHDS, are important for the results.

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 as, childhood IQ, early childhood family resources, association with deviant peers at age 15 and the proportion of the individual's 10^{th} grade who go on to the 11^{th} 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. This result has implications for the analysis of academic performance, showing that, even in the absence of a number of relevant variables, the use of selection correction methods does not change the results much.

We have shown that the inclusion of personal, family and teenage peer information has the potential to influence the conclusions drawn from studies. For example, the results from the model indicate that keeping students at school would only be a first step in improving the education level of the student population, since the academic performance of those who would have left without the policy is likely to be poor. The model with a limited number of variables did not show so clearly that more needs to be done in terms of policies than preventing students from leaving school. In addition, it was shown that (the often unobserved) school and peer characteristics contribute to academic attainment. If these variables cannot be included, variables such as school ethnicity (Rothstein, 2004) may serve as a proxy for these factors. School and peer characteristics may be influenced through policy on learning and behavioural aspects of teenagers' school and peer environments, as opposed to ethnicity itself. Therefore, identifying these underlying factors is important to enable appropriate policy design.

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^{th} 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).
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.
Birth Month_ Semester 1	Binary=1 if born in April or May (school semester 1 in the southern hemisphere), =0 if born in June, July or August (school semester 2).
Mother without Qualifications (Mother_No_Qualifications)	Binary=1 if child's mother does not have formal educational qualifications (10^{th} 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^{th} 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.
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.

Table A1: continued

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.[17]
Registered Unemployment (Local_Unemployment)	Regional unemployment rate by gender in which each individual was living at 15 years of age. [18] 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). This test is conducted by a trained Psychologist. (IQ score was not reported to the child, the parents or teachers).
Class Size (Ave_Class Size)	Average class size in secondary school
Proportion of Students Continuing (Proportion of Class_Continue)	Proportion of an individual's 10^{th} grade (Fifth Form) class within the data set continuing onto the 11^{th} grade. The relevant individual is excluded from the calculation.
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.

17 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.

18 Source of this information is the 1991 Census of Population and Dwellings: Regional Summary.

Table A2: Childhood IQ, Family Income and Grade Characteristics by Birth Month Sample

Characteristics	**Mean** (Standard Deviation) **Birth Month Falls in:** Semester 1	Semester 2
Personal Characteristics		
IQ (tested at 8 years of age)	103.9 (13.97)	103.4 (14.79)
Average School Certificate Grade Point Average (age 15 and 10th grade, where Fail=0, C=1, B=2, A=3)	1.08 (0.78)	1.10 (0.88)
Adolescent Average Income Decile: Ages 11-14 (10 is most affluent decile)	5.60 (2.55)	5.68 (2.60)
Early Childhood Average Income Decile: Ages 1-5	5.83 (2.40)	5.94 (2.40)
Own their Home	89.0%	89.2%
Left school before the exam	7.5%	11.0%
Sample Size:	252	409

Bibliography

Angrist, JD and Krueger, AB (1992), "The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments From Two Samples", *Journal of the American Statistical Association*, 87 (418), 328-336.

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.

Card, D and Rothstein, J (2005), "Racial Segregation and the Black-white Test Score Gap", *NBER Working Paper,* March.

Carneiro, P, and Heckman, JJ (2002), "The Evidence on Credit Constraints in Post-secondary Schooling", *The Economic Journal*, 112, 705-734.

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.

Heckman, JJ and Krueger, A (2003), eds., *Inequality in America: What Role for Human Capital Policy?*, Cambridge, MA: MIT Press.

Maani, SA (1997), *Investing in Minds: The Economics of Higher Education in New Zealand,* Wellington, Publisher: Institute of Policy Studies.

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*, Elsevier, 217-248.

Maani, SA 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.

Micklewright, J (1989), "Choice at Sixteen", *Economica*, 56, 25-39.

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.

Rothstein, JM (2004), "College Performance Predictions and the SAT", *Journal of Econometrics,* 121, 297-317.

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.