YOUTH IN FOCUS PROJECT DISCUSSION PAPER SERIES
No. 4, December 2008

Exploring the Factors Associated with Youths' Educational Outcomes: The Role of Locus of Control and Parental Socio-Economic Background

Juan D. Barón

youthinfocus.anu.edu.au

Jointly funded by the Australian Government Department of Families, Community Services and Indigenous Affairs and the Australian Research Council and carried out by The Australian National University

Youth in Focus Project Discussion Paper Series
ISSN 1835-4025

Youth in Focus
Research School of Social Sciences
ANU College of Arts and Social Sciences
Coombs Building 9
The Australian National University
Canberra ACT 0200

Exploring the Factors Associated with Youths' Educational Outcomes: The Role of Locus of Control and Parental Socio-Economic Background*

Juan David Barón R.†
Research School of Social Sciences—Economics Program
Australian National University

Version: 2.5.0

Abstract

Using unique information for a cohort of Australian youth, this paper explores the association between youths' perception of control (i.e. locus of control) and three educational outcomes: (i) Year 12 completion, (ii) whether youth obtained an Equivalent National Tertiary Entrance Rank (ENTER) score, and (iii) the actual ENTER score. By using a measure of socio-economic status based on 12 years of parental income support histories, the paper also investigates the association between growing up in a socio-economically disadvantaged household and subsequent educational outcomes. Additionally, the paper considers the hypothesis that disadvantage has an indirect effect on youths' educational outcomes through its effect on locus of control. The results suggest that youths with a more internal locus of control (e.g. those who believe their actions determine their future outcomes) are more likely to complete Year 12, more likely to obtain an ENTER score, and obtain better ENTER scores. The evidence is also consistent with a negative relationship between disadvantage when growing up and youths' educational outcomes. Even after controlling for demographic and family characteristics, youths who grew up in socio-economically disadvantaged households are up to 10 per cent less likely to complete Year 12 and up to 20 per cent less likely to obtain an ENTER score. There is however no evidence of an indirect effect of being disadvantaged on educational outcomes through the effect of disadvantage on locus of control once other characteristics are accounted for. Although highly disadvantaged youths obtain ENTER scores that are four points lower than those of non-disadvantaged youth, locus of control shows only a small association with actual ENTER scores.

JEL-Classification: I38, J24, H31
Keywords: locus of control, parental socio-economic background, education

*The data used for this research come from the Youth in Focus Project which is jointly funded by the Australian Government Department of Education, Employment and Workplace Relations, the Australian Government Department of Families, Housing, Community Services and Indigenous Affairs, Centrelink and the Australian Research Council (Grant Number LP0347164) and carried out by the Australian National University. I also want to thank Deborah Cobb-Clark, David Ribar, Chris Ryan, Michelle Tan and participants at the IZA internal seminar for useful comments and discussions.

†PhD Scholar, RSSS, Economic Program. Address: Coombs Building (Bldg 9), Research School of Social Sciences, The Australian National University, Canberra, ACT 0200, Australia. Phone: (61 2) 6125 8109. Fax: 61 2 6125 0182. E-mail: Juan.Baron@anu.edu.au.

1 Introduction

In standard economic models of human capital accumulation, individuals decide to invest in education when the future benefits of such decisions outweigh the benefits of alternative decisions (Becker, 1993). The benefits individuals use in this exercise will implicitly and inevitably reflect their expectations, either optimistic or pessimistic, about the future returns to investments in education. This reasoning opens the possibility for factors such as personality to have a potential role in explaining individuals' educational outcomes through their effects upon their expectations about future outcomes. The present paper looks at one such personality characteristic called Locus of Control. In brief, locus of control is a psychological concept that measures people's perception of their control over the things that happen to them. In other words, it measures whether people believe their actions affect what happens to them in the future. Those who believe success comes from hard work are described as having a more *Internal* locus of control whilst those who believe success and failure are random events independent of the effort they put in are described as having a more *External* locus of control. Despite the interest of the profession in the relationship between personality and educational outcomes, evidence supporting the existence of the relationship between them is scarce in the Australian context.

In a different vein, previous research in Australia suggests that children who grow up in socio-economically disadvantaged households are significantly more likely to drop out of school and are under-represented at universities. Given the lack of information on parental income when children are growing up, in most of these studies, the measure of parental socio-economic status is based on parental occupation at a single point in the past. Using a recent and unique dataset for a six-month cohort of Australian youth, this paper also studies the relation between parental socio-economic background and youths' educational outcomes. One of the contributions of this paper is to provide alternative estimates of this relationship using instead almost 12 years of information on parental income support use. With these data it is also possible to partially investigate whether the relationship between disadvantage and educational outcomes varies by youth's age at exposure to disadvantage. This is possible because the measure of dis-

advantage, or socio-economic background, covers most of youths' lives (approximately four to sixteen years old).

The connection between this paper's interest in the relation between locus of control and educational outcomes, on one hand, and the same outcomes and parental socio-economic background (i.e. level of disadvantage when growing up), on the other, stems from previous findings in the literature that suggest that income support use affects certain personality characteristics such as self-esteem or attitudes towards work. Little is known, however, about the effect of parental income support use (i.e. parental background or disadvantage) on the personality and attitudes of their children—attributes that might well affect their educational outcomes. This has important policy implications because if parental disadvantage indeed affects personality characteristics that influence educational outcomes, then disadvantage (as well as income support reliance) becomes a vicious cycle that gets transmitted from parents to children. The present paper investigates this hypothesis by regressing locus of control on the measures of parental disadvantage.

This paper is interested in the following questions: Is locus of control related to educational outcomes? Is there evidence of a relationship between parental socio-economic status, or growing up disadvantaged, and youths' educational outcomes? Is there any evidence of a relationship between locus of control and disadvantage that might suggest an indirect effect of disadvantage on youths' educational outcomes through locus of control? Lastly, does the relationship between disadvantage and youths' educational outcomes depends on how old youths were at exposure to disadvantage?

The analysis uses three educational outcomes: whether the youths (i) completed Year 12, (ii) obtained an Equivalent National Tertiary Entrance Rank (ENTER) score, and (iii) the actual ENTER score. The idea behind using these outcomes is to capture not only the youths' decisions regarding school but also their performance. There are at least two main challenges in working with these educational outcomes and locus of control. The first one stems from the fact that there are several measures of locus of control available in the data. When confronted with this situation the traditional approach in the economics literature is to combine these by adding them up and standardising the resulting variable to have mean zero and variance one. Unfortunately, results might be sensible to this *ad*

hoc strategy. In the approach taken in this paper, these measures of locus of control are combined in a more flexible way by including an extra set of equations that relate the latent variable to each of the observed measures. A clear advantage of this approach is that assumptions are made explicitly. The second challenge in using these educational outcomes is sample selection, particularly due to "incidental truncation". The problem in the Obtained ENTER Score model, for example, is that people who finish school might not be representative of the population of Year 12 students. If this is the case, and selection is not dealt with, the Obtained ENTER Score model will be estimated using a non-random sample of the population rendering biased estimates. This paper deals with this selection problem by using as an instrument the finding that youths who were born early in the sample are more likely to graduate earlier than other youths. This is significant since the data only include a six-month cohort of Australian youth (start October 1987 and finish in March 1988). When I analyse the ENTER score itself, the results are conditional on youths obtaining one.

The Youth in Focus Project provides the data for the analysis. The project interviewed a six-month cohort of approximately 4,000 Australian Youth (aged 18 at the time of the interview) about their educational outcomes, locus of control, and demographic and family characteristics. These survey data are complemented by administrative information on almost twelve years of parental income support history for these youths' families.

Consistent with the international literature on locus of control and educational outcomes, the findings support the hypothesis that young people with a more internal locus of control have a higher probability of finishing Year 12 and obtaining an ENTER score. There is evidence of a small but statistically significant relationship between locus of control and ENTER scores, conditional on getting an ENTER score.

The results in this paper also suggest a negative relation between growing up disadvantaged and youths' educational outcomes. There is however no evidence of a relationship between growing up disadvantaged and subsequent locus of control. An effect of disadvantage on youths' educational outcomes through locus of control is hence not borne out by the data. The results also suggest that compared to those growing up in non-disadvantaged families, children who grow up in disadvantaged families are up

to 10 per cent more likely either to finish school late or drop out and up to 20 per cent less likely to obtain an ENTER score. Since finishing school and obtaining an ENTER score are the first steps toward attending university, it is highly likely that disadvantaged youth in the sample will be under-represented in tertiary education.

Conditional on being disadvantaged for less than six years, there is no evidence of a differential effect of disadvantage on youths' educational outcomes depending on the time at which youth were exposed to disadvantage. Finally, although highly disadvantaged youths obtain ENTER scores that are on average four points lower than those of non-disadvantaged youth, locus of control shows only a small, but statistically significant, association with obtaining ENTER scores. Conditional on school completion and obtaining an ENTER score, individuals with a more internal locus of control obtain slightly better ENTER scores.

The outline of this paper is as follows. By means of a brief literature review, the next section sets up the conceptual framework for the relationship between educational outcomes and the explanatory variables of interest (i.e. locus of control and parental socio-economic background). Section 3 introduces the dataset and gives some basic statistics. Section 4 describes the econometric model. By allowing each of these indicators to be determined by the latent locus of control, the model used in this paper improves on the econometric treatment of latent variables when there are (multiple) imperfect measures for them. Section 5 presents estimation results and Section 6 discusses some extensions. Section 7 concludes this paper.

2 Literature review

Economists have for some time been interested in the effect of personality characteristics, or non-cognitive skills, on educational and labor market outcomes (Goldsmith et al., 1997; Heckman et al., 2006; Osborne-Groves, 2005).[1] Given this interest, concepts such as self-esteem, pessimism, initiative and locus of control, among others, are now common in the jargon of labour economists. Osborne-Groves (2005), for example, finds that

[1]Other social scientists have, for a long time, examined the relationship between personality and some labour and educational outcomes. Some of the applied work in this literature however is based on convenience samples which are small and unrepresentative.

a variety of personality measures have a statistically significant association with wage levels, even after controlling for measures of cognitive ability. Controlling for ability in wage regressions is important because if, as economists believe, ability is correlated with non-cognitive measures included in the regression and ability itself is not included, estimates of the effect of non-cognitive characteristics on wages will be biased. Other authors have also shown that even when taking into account econometric problems such as measurement error in the non-cognitive measures and reverse causality, non-cognitive skills affect wages and schooling and, in addition, are associated with risky behaviors such as marijuana use, imprisonment, and illegal activities (Heckman et al., 2006). Studies like these are representative of a growing literature providing evidence of the relationship between personality measures, education and labor market outcomes (see Bowles et al., 2001). Evidence in the Australian case is nevertheless scarce.

One of the most widely used personality measures in economics literature is Rotter's Locus of Control Scale (Rotter, 1972, 1990).[2] One reason for this is the availability of the measure in large socio-economic surveys such as the National Longitudinal Survey of Youth (NLSY) and the Panel Study of Income Dynamics (PSID) in the United States; the National Child Development Survey in Britain; or the Household, Income and Labour Dynamics in Australia (HILDA). More importantly, the concept is intrinsically appealing to economists. Locus of control refers to the way people see themselves in control of the events that happen to them, and the power they have to change them. The concept categorises individuals into one of two groups: those who believe that good things happen to them because they work hard (*internal* locus of control) and those who believe that what happens to them is the product of luck or destiny (*external* locus of control). The internal versus external distinction makes locus of control an attractive concept to explain why some people finish school and others do not, or why some have greater success in the labour market. In this paper I assume (as do psychologists) that locus of control is a continuous measure where internals are at one end of the spectrum and externals at the other.

[2]Rotter's original locus of control scale was based on 29 questions that allowed only two contrasting alternatives as answers (Rotter, 1972). Some other measures that allowed differing degrees of agreement with the statements and fewer questions have also been developed. The analysis in this paper uses the Pearlin and Schooler (1978) Mastery Scale.

Having realised the potential of locus of control to explain some of the variation in wages and educational attainment, economists started to look for evidence of the relationship between locus of control and wages, on one hand, and locus of control and educational outcomes on the other.[3] The empirical support for the relationship between locus of control and outcomes is at best mixed. Whilst some authors find strong evidence of a relationship with education levels and labour market outcomes, such as years of schooling, occupational advancement, and earnings, (Andrisani and Nestel, 1976; Andrisani, 1977; Coleman and DeLeire, 2003; Osborne-Groves, 2005; Goldsmith et al., 1997), others reject these claims, sometimes using similar datasets (Cebi, 2007; Duncan and Morgan, 1981; Murnane et al., 2001). In addition, Linz and Semykina (2008) find no evidence of a relationship between locus of control and performance at work. Despite these mixed results, researchers do find a consistent association between locus of control and behaviours such as daily smoking, drug use, truancy, and involvement in crime early in life (Carneiro et al., 2007; Clarke et al., 1982; Heckman et al., 2006).

Many studies in this literature treat locus of control as a fully observed and perfectly measured variable. They overlook the fact that survey information provides only imperfect measures of locus of control.[4] The approach often taken in economics literature is to aggregate the multiple measures of locus of control found in surveys into a single index and then use conventional regression techniques. The weights underpinning the index, however, would necessarily be *ad hoc*, given the lack of information about the contribution that each measure makes in predicting an individual's locus of control. Unfortunately, estimation results are likely to be sensitive to the weights chosen. The methodology used in this paper allows more flexibility in the way the different measures are combined (see Section 4).

Despite the numerous attempts to empirically document the relationship between locus of control and educational outcomes, only a few researchers in economics have put forward hypotheses about the mechanism for the existence of this relationship. Andrisani (1977) argues, for example, that since internals believe success comes from hard work, they should be more likely to be aware of information that could be used for future

[3] Earlier research by psychologists suggests that these correlations can be observed.

[4] Papers that consider the latent nature of locus of control or other psychological concepts in their methodologies are Osborne-Groves (2005) and Heckman et al. (2006) among a few others.

decision-making; are more willing to take action to improve; and are less likely to surrender to peer pressures. More importantly, Andrisani argues, internals are more likely to acquire lager amounts of valuable human capital. Under his conceptual framework, locus of control affects wages through motivation and initiative and also through the effect of these factors on past schooling decisions. Coleman and DeLeire (2003), more than twenty years later, incorporated these ideas in an explicit model of human capital accumulation. In their model, locus of control influences young people's perceptions about the future returns to education. Since internals believe their actions regarding education will have a large impact on their future outcomes, they will tend to accumulate more human capital than individuals with a more external locus of control. Using data from the National Educational Longitudinal Study (NELS) in the United States, Coleman and DeLeire (2003) provide evidence of the validity of their model. More recently, however, Cebi (2007) found no evidence of the implications of this model using data from the NLSY in the United States.[5]

Another branch of the literature that relates to this paper is the one discussing the role of parental background (i.e. parental socio-economic status) on the educational outcomes of children. For a review of the literature on the relation between family background and educational outcomes in the United States see Haveman and Wolfe (1995). In Australia, the studies by Marks (2004), Le and Miller (2005), and Cardak and Ryan (2006) study the relationship between educational attainment and parental socio-economic status. In most Australian studies the measure of parental socio-economic background is based on an index created using parental occupation earlier in the life of the youth. The present paper uses 12 years of parental income support history to define an alternative measure of parental socio-economic background (i.e. the youth's level of disadvantage at home).

The results from this literature that are of most interest are: 1) children who grow up in socio-economically disadvantaged families tend to have lower educational attainment than comparable children from non-disadvantaged families; 2) growing up

[5]In this literature, economists worry about the potential endogeneity of locus control in human capital accumulation and wage equations. If ability is just a proxy for locus of control, and it is not controlled for in these regression equations, the estimate of the effect of locus of control on educational outcomes or wages is biased.

in single-parent families also has detrimental effects on educational outcomes; and 3) youths from disadvantaged households are under-represented in tertiary education. The present paper evaluates the first two of these hypotheses for the case of Australia, and provides some insights into the third one by looking at the decision to obtain an ENTER score (i.e. in general, a prerequisite to attend university).

In a relevant study, Ku and Plotnick (2003) investigate in more detail the relationship between parental income support use—welfare, in the United States— and children's outcomes. By using data from the PSID, the authors find that mother's income support use has a detrimental effect on their children's educational attainment, particularly so when the children are exposed to the income support system during late childhood and adolescence. This study also documents that (except for exposure during early childhood) cross-sectional estimates do not seem to differ from those obtained when eliminating unobserved time-invariant family heterogeneity—sibling to sibling differences are used to eliminate the family fixed effect. Ku and Plotnick suggest that if any difference exists between cross-section and panel data estimates for this relationship, that difference indicates that eliminating family fixed effects makes the effect of income support on educational attainment appear larger in magnitude. This implies that cross-section estimates might be biased towards zero (see Ku and Plotnick, 2003, Table 4).[6] The authors find that compared to youth not exposed to income support, youth from income support families are 24 per cent more likely to drop out of school.

So far this review has briefly discussed the literature that associates locus of control and the educational outcomes of young adults. It has also summarised the main literature findings on the relationship between parental socio-economic background and youths' educational outcomes. But, how do these two branches of the literature relate to each other for the objectives of the analysis? Gottschalk (2005) and Elliott (1996) argue that income support use (i.e. disadvantage) might alter the attitudes, beliefs, and even personality characteristics of recipients. Income support use might reduce the stigma of relying on public funding or change people's self-esteem. It is unknown, however, whether parental income support use, or disadvantage in general, affects the views or personality characteristics of children growing up in this environment. If disadvantage at

[6]Other social scientist have also studied this issue, see for example Rich (1999).

home affects youths' personality (e.g. locus of control) there might be a potential indirect effect of disadvantage upon educational outcomes through the impact of disadvantage on personality. To evaluate the plausibility of this hypothesis, the empirical part of this paper regresses (latent) locus of control on some family characteristics and the measures of disadvantage.

This paper aims at contributing to both the literature that relates personality (i.e. locus of control) and youths' educational outcomes, on one hand, and parental socio-economic background and educational outcomes on the other. In the process, the econometric model used is novel in the way it treats the available measures of the latent locus of control variable.

3 Data

3.1 The Youth in Focus Data

The analysis uses data from the Youth in Focus Project (YIF).[7] The YIF data are unique in providing for a sample of 18-year-old youths detailed information about education (e.g. not only Year 12 completion but also ENTER scores), parental income support histories, and individual and household characteristics.[8]

Specifically, the YIF Project uses Australian administrative social security records to identify all young people born in the six-month period between October 1987 and March 1988 who ever had contact with the social security system between 1993 and 2005 (see Breunig et al., 2007). These social security records provide high-quality, fortnightly data on the payment details for the universe of Australians receiving a wide range of social benefits. Although young people can appear in the administrative data if they receive social security payments themselves, most enter the system because a family member (generally a parent) received a payment which depended in part on the youth's relationship to the payee.[9] At some point many families received a benefit that is best thought

[7] For more information about the project see http://youthinfocus.anu.edu.au.

[8] Mother were also interviewed, but their information is ignored in the analysis in this paper. The reason is that I can only find mother's information for 60 per cent of the youth. If I use mother's information, my sample of youth will reduce by approximately 40 per cent.

[9] In general, youths can start receiving income support payments in their own right from the time they turn 16 years old.

of as income support, e.g., unemployment benefits or sole parent payments; however, many others did not. Approximately 40 per cent of families in the administrative data received only family tax or child care benefits during the period covered by the data.[10] Given the generosity of the Australian social security system, the YIF's research team estimates that approximately 90 per cent of young people in the relevant six-month birth cohort are in the administrative data.[11] The YIF project summarises a family's income support history by using the administrative data to categorise youths and their parents into one of six groups, depending on the recency and intensity of the family's income support receipt.[12] Specifically, families who received income support payments for a total of six years or more (out of 12 possible years) are classified as having had an intensive exposure to income support. At the other end of the spectrum are families that received no income support benefits at all. In between, are roughly 30 per cent of families who had a more limited exposure to the income support system at some point in the previous 12 years. A stratified random sample of young people and the corresponding parent or guardian—in 96.5 per cent of cases the biological mother—was selected for interview from the administrative data. In order to ensure adequate samples of income support recipients for analysis, the stratification into six groups was done on the basis of intensity and recentness of income support receipt (see Breunig et al., 2007). Data from separate phone interviews with the youth and their parents, as well as a self-completion questionnaire administered to youth, were then matched to the administrative social security data.[13] The analysis includes youth matched to the administrative income support data of their parents. Including survey information from the Parent's Questionnaire substantially reduces the sample of youth; this information, therefore, is not included in the

[10]The Family Tax Benefit is essentially an income tax credit to families with children rather than a welfare payment. Currently (2008) a family with two children would receive a Family Tax Benefit for incomes up to $105,000 AUD.

[11]In particular, the Australian social security system is nearly universal, with some benefits, e.g., Child Care Benefit, having no income test, and other benefits such as Family Tax Benefit being denied only to those household in the top 20 per cent of the income distribution. Comparing the YIF youth sample with Australian Census data suggests that the administrative data capture roughly 90 per cent of the youth born in the period (Breunig et al., 2007). See Centrelink (2007) for more information about the Australian social security system.

[12]Appendix Table A1 describes the stratification categories in the YIF data. With some modifications, the measure of disadvantage used in the present paper is based on these categories.

[13]The survey response rate was 34.2 per cent for parents, and 34.7 per cent for youth—73.1 per cent of whom also completed the self-completion questionnaire. More than 96 per cent of young people and 92 per cent of parents completing the survey consented to having this information linked to their administrative records.

present analysis.

The analysis necessarily restricts the sample in a number of ways: first of all, some of the questions used in the analysis come from the Youths' Self-Completion Questionnaire (SCQ). This means that the analysis is subject to a smaller response rate attributable to the use of variables in the SCQs.[14] The main questions in the SCQ included in the analysis are the locus of control questions and two questions on non-pecuniary parental investments in youths' education (i.e. whether parents read to their child at night, and whether they helped with homework). There are approximately 1,150 youths with missing information in the SQC relevant variables. Additionally, there are approximately 750 youths who did not provide information on some of the other explanatory variables (note: most of these are due to missing parental education information for either mother or father from the youth questionnaire). Consequently, the estimation sample consists of approximately 2,100 youths who have complete survey information for the variables of interest. Appendix Table A2 presents summary statistics of the variables in the analysis.

3.2 Educational outcomes, locus of control and parental background

The analysis considers three educational outcomes: (i) Year 12 completion, (ii) whether the individual obtained an Equivalent National Tertiary Entrance Rank (ENTER) score and, (iii) conditional on obtaining it, the ENTER score itself. The Year 12 completion variable is a dichotomous variable taking value 1 for a youth who, at the time of the interview, had finished Year 12 or equivalent.[15] The second educational outcome is also a dichotomous variable that takes value 1 if the youth, at the time of the survey, had obtained an ENTER score, and 0 otherwise.

The third and final educational outcome is the actual ENTER score. Because individuals who obtained ENTER scores in the bottom 30 per cent of the distribution are recoded to 30, the variable used in the analysis takes values between 30 and 99.9.[16] To

[14] It is well-known that self-completion questionnaires have lower response rates than face-to-face or telephone interviews. The YIF was not an exception, although the payment of an incentive increased response rates, particularly for disadvantaged populations.

[15] Youth were interviewed between August and December 2006. At that time youths were at least 18 years old. Under normal circumstances Australian youth finish Year 12 at age 18, and obtain the ENTER score at the end of Year 12. Coding the Year 12 completion "at the time of the interview" is standard practice. See, for example, Evans and Schwab (1995) and Altonji et al. (2005).

[16] Admission to university based on ENTER scores is the most common form of admission to tertiary education in Australia (Cardak and Ryan, 2006). To obtain an ENTER score, youth complete a pre-university

account for "institutional censoring", the model used for this outcome is a censored regression model.

An alternative educational outcome is whether youths were enrolled at university at the time of the survey. The YIF survey however interviewed youths too soon after completing Year 12 for this to be an appropriate outcome to consider.

The survey asks young people about their feelings when facing problems and their perception of control over their lives and the things that happen to them. In particular the survey asked respondents whether they Strongly Agree, Agree, Disagree, or Strongly Disagree with seven statements about 1) solving their problems, 2) feeling pushed around in life, 3) controlling things that happen to them, 4) achieving self-imposed goals, 5) feeling helpless when dealing with problems, 6) controlling things that happen to them, and 7) having the power to change things in their lives. Responses from these seven questions constitute the basis of the measure of locus of control in the analysis.[17]

Table 1 presents raw evidence of the association between locus of control and the three educational outcomes under analysis. To facilitate the interpretation, for each locus of control variable an indicator, taking value one if the individual Strongly Agrees or Agrees with the statement, is created and zero otherwise. The figures in the Year 12 completion panel (the first educational outcome) are interpreted as follows: youths who agree they cannot solve some of the problems they have are significantly less likely

entrance program in Year 12. The result obtained is based on state-wide tasks and examinations; and reflects the percentile rank of each individual's performance within the cohort. The name and scope of ENTER scores vary across Australian States (see Marks et al. (2001) for details).

[17]The exact wording of the question is: "The following statements describe the way some people feel about how much control they have over their lives. How strongly do you agree or disagree (*Strongly disagree*, *Disagree*, *Agree*, and *Strongly Agree*) with the following statements? (i) There is really no way I can solve some of the problems I have; (ii) Sometimes I feel that I'm being pushed around in life; (iii) I have little control over the things that happen to me; (iv) I can do just about anything I really set my mind to; (v) I often feel helpless in dealing with the problems of life; (vi) What happens to me in the future mostly depends on me; and (vii) There is little I can do to change many of the important things in my life." This question is called the Pearlin and Schooler's (1978) Mastery scale. In economics literature most researchers use simplified measures based upon Rotter's (1972) locus of control scale. See, for example, Heckman et al. (2006), Coleman and DeLeire (2003), Cebi (2007). The original Rotter's locus of control scale has been criticised for forcing the interviewee to choose between two supposedly extreme answers. Ray (1984) argues that the locus of control scales obtained in this way have no internal validity because many respondents answer that both statements are applicable to them. More recent instruments, such as the one used here, allow for different degrees of agreement or disagreement with the statement and are designed to overcome this criticism. A longitudinal survey, the Household, Income and Labour Dynamics in Australia (HILDA) Survey also includes the Pearlin and Schooler's (1978) Mastery Scale in the 2003 and 2004 Waves. The present analysis does not use HILDA data because samples of youth are smaller.

to have finished Year 12 at the time of the survey by almost six percentage points (18.3 versus 24.6 per cent). Those feeling pushed around in life are less likely to have graduated although the difference is not statistically significant. People who think they have little control over the things that happen to them are also more likely to have not completed Year 12 (19.7 per cent); only 13.9 per cent of Year 12 graduates share this view. Rather, individuals who think they can do anything they set their minds to are significantly more likely to finish school. Dropouts and Year 12 graduates appear to differ the most in how they feel in dealing with their problems, with dropouts being significantly more likely than Year 12 graduates (37.6 versus 28.6 per cent) to agree with the statement that they feel helpless in dealing with their own problems. Finally, believing that what happens to you depends upon yourself makes individuals in the sample only 2.2 percentage points more likely to complete Year 12 (a difference that is statistically significant). Despite these views, people who say they can do little to change things in their lives are almost 6.4 per cent less likely to complete Year 12.

[Table 1 here]

Table 1 also presents similar statistics for the second educational outcome, obtaining an ENTER score. The figures in this panel are conditional on having completed Year 12. Youth who feel they cannot solve their problems along with those who feel pushed around in life and those feeling helpless in dealing with their problems are all less likely to obtain an ENTER score by statistically small and economically insignificant margins. In contrast, people who agree they have little control over the things that happen to them and those who feel there is little they can do to change things are significantly less likely to obtain an ENTER score by almost six and nine percentage points respectively. Those who think they can do anything are also seven percentage points more likely to obtain an ENTER score (70 versus 62.6 per cent). In general, people with a more internal locus of control seem more inclined towards obtaining ENTER scores. These figures might be subject to some bias due to the potential non-randomness of the sample of youths that completed school. The econometric model for this outcome, described in the next section, attempts to deal with the potential selection problem due to this "incidental truncation".

The last panel of the table presents the average ENTER score for those who Agree

or Strongly Agree with each of the locus of control statements, and for those who Disagree or Strongly Disagree with them. The difference in average ENTER scores for these two groups is in general small in size and statistically insignificant (less than 1 percentage point for almost all statements). Youths who think they have little control over the things that happen to them, nevertheless, obtain ENTER scores that are on average 3 percentage points lower. This suggests that externals obtain on average lower ENTER scores (conditional on having finished school and having obtained an ENTER score).

All in all, the figures in Table 1 suggest that youths who exhibit a more internal locus of control tend to complete Year 12 and obtain ENTER scores in higher rates than externals. Internals also obtain higher ENTER scores, albeit the evidence being weaker for this outcome. The strength of these relationships, nevertheless, varies according to the locus of control measure under consideration. This highlights the challenges researchers face when considering how to combine these locus of control variables. The standard practice in the literature is to either add the different answers for each individual and then standardise the resulting variable to have mean zero and variance one, or carry the analysis using one variable at a time. In contrast, the econometric models presented below assume that locus of control is a latent variable and the information in the survey contains only imperfect measures of it. This is done by incorporating a measurement model for locus of control in the main educational outcome regression and estimating the system jointly. The advantage of this approach is the imposition of less *ad hoc* restrictions on the way (imperfect) measures of locus of control are combined, in addition to potential gains in the precision of the estimates. Importantly, the assumptions are stated explicitly.

Unlike previous research in Australia on the effect of parental socio-economic status on the educational attainment of youth (Cardak and Ryan, 2006; Marks et al., 2000; Le and Miller, 2005), the measure of disadvantage (e.g. parental socio-economic status) used in the present analysis is based on 12 years of parental income support history and not on parental occupation at some point in the past. This information is provided by the stratification variable in the YIF data described in the data section. As described there, this measure contains six categories based on recentness and duration of parental income support use. Appendix Figure A1 reports the stratification categories. The category "no-history of income support" is not included. From the figure it is clear that to be able

to evaluate whether the relationship between disadvantage and educational outcomes changes with the age at which youths were exposed to disadvantage, it is necessary to combine Strata D and F. This will allow the comparison of the columns in the figure; interpretation, however, is conditional on being on parental income support (i.e. exposure to disadvantage) of less than six years because Strata B cannot be further divided.

Table 2 presents means and standard deviations of youths' educational outcomes by different degrees of parental socio-economic disadvantage. The statistics in the left panel of the table suggest a strong association between growing up in a family exposed to disadvantage and the probability that the youth will not graduate from school. Specifically, while the probability of completing Year 12 for youth who grew up unexposed to disadvantage is 77.6 per cent, the probability is almost 23 percentage points lower (54.9 per cent) for youth that grew up in heavily disadvantaged families (those in which parents received income support for more than six years). This difference is statistically different from zero conventional levels as implied by the p-value.[18] Youths in intermediate categories of disadvantaged families (those in which parents received income support for less than six years) also show significantly lower probabilities of completing Year 12 when compared to non-disadvantaged youth. For these groups the probability ranges from 64 to 69 per cent. Youths first exposed when they were older than six years of age are approximately four percentage points more likely to have completed Year 12 than those exposed at other ages, suggesting that being disadvantaged early in life has a slightly bigger effect on educational outcomes.

The results are qualitatively and quantitatively similar whether or not an ENTER score was obtained. Conditional on Year 12 completion, non-disadvantaged youth are 23 percentage points more likely to obtain an ENTER score than the most disadvantaged youth (76.8 versus 53.7 per cent). Youth experiencing disadvantage for less than six years have significant lower probabilities of obtaining an ENTER score, by between 8 and 14 percentage points, than the non-disadvantaged group (but higher probabilities than the highly disadvantaged group). In addition, the younger the youth at the time of exposure to disadvantage at home, the lower the probability of completing Year 12. This paper

[18]The interpretation of the results are in terms of "disadvantage" rather than parental income support exposure because there are no measures of parental income at the time the youth was growing up.

explores this timing-of-exposure issue in the extensions section below.

[Table 2 here]

Table 2 also presents evidence of considerable differences in ENTER scores when comparing results by youths' experience of disadvantage. Whilst non-disadvantaged youth (Stratum A) obtain on average 75.06; intermediately disadvantaged youths obtain lower scores (between 71.64 and 72.65 in Strata C, D and F, and E). The largest difference, however, in relative performance is for highly disadvantaged youth (Stratum B): they obtain scores almost 5.6 points lower than non-disadvantaged youth.

The remaining question is whether parental socio-economic background associates with youth's locus of control in a way that affects their educational outcomes. For this to be true, it must be the case that the measures of locus of control and measures of disadvantage exhibit some degree of association. To explore this, Table 3 presents the proportion of highly disadvantaged youth (Stratum B) and non-disadvantaged youth (Stratum A) who Agree or Strongly Agree with the locus of control statements. Highly disadvantaged youth are almost five percentage points more likely to agree they cannot solve some of the problems they have. With small variation in the differences, the results for the other six statements is consistent with the hypothesis that youth from disadvantaged families are more likely to develop an external locus of control. Differences are between 2.1 and 4.5 percentage points. It is then plausible, at least when not taking anything else into account, that growing up in a disadvantaged household has an indirect effect on educational outcomes through its negative effect on locus of control. This hypothesis is formally studied in the extensions section by running a model of youth's locus of control on disadvantage measures while controlling for other factors.

[Table 3 here]

Taken together the simple associations in these tables suggest that youths' perceptions of control, whether they are internal or external, and their socio-economic disadvantage when growing up might potentially affect educational outcomes (both directly and indirectly through locus of control). Moreover, there is also a potential difference in youth's outcomes due to their exposure to disadvantage depending upon their age. The

following sections analyse the robustness of these relationships to the inclusion of other covariates and by taking into account the latent nature of locus of control.

4 The econometric model

One of the primary empirical challenges is to make the best use of the multiple variables of each individual's locus of control. In this situation, the approach the economics literature often takes is to aggregate the multiple measures of locus of control into a single index and then use conventional regression techniques. In the present case, however, the weights underpinning the index would necessarily be *ad hoc* given the lack of information about the contribution that each makes in predicting the individual's locus of control (summing them, for example, assigns each an equal weight). Unfortunately, estimation results are likely to be sensitive to the weights chosen. Alternatively, other researchers in similar situations prefer to analyse each measure separately (Dohmen et al., 2006 follow this approach to analyse different measures of trust). The difficulty with this single-equation, measure by measure approach, is that it treats the data as though each survey question provides information about a separate, perfectly measured concept. Instead, the models described below allow for the possibility that answers to specific survey questions are only imperfect measures, or indicators, of a single concept called locus of control. Additionally, combining the information from several measures may improve the precision of the estimates. Consequently, the model for each educational outcome consists of two parts. The first is the main equation for the outcome of interest that contains as a covariate the latent locus of control. The second part is a measurement model which relates the observed (ordered) variables, or indicators, to the underlying latent variable. The next three subsections describe the specifics for each of the three educational outcomes.

4.1 A model of high-school graduation with latent locus of control

Define the propensity of completing Year 12 as y^*, a latent variable, such that

$$y^* = \mathbf{X}\boldsymbol{\beta} + \gamma LC^* + \mathbf{W}\boldsymbol{\theta} + u; \tag{1}$$

where $\mathbf{X}$ represents a set of covariates, LC^* is locus of control, and $\mathbf{W}$ is the set of dummy variables indicating youths' exposure to disadvantage at home. $\{\boldsymbol{\beta}, \gamma, \boldsymbol{\theta}\}$ are vectors of parameters of conformable dimensions to the variables they multiply, and u is the error term which is independent of each element of $\mathbf{Z} = \{\mathbf{X}, LC^*, \mathbf{W}\}$. In addition, $u \sim N(0, \sigma_u^2)$. The researcher cannot observe y^*. Instead the researcher observes y taking values 1 and 0 according to the rule $y = \mathbf{1}[y^* > 0]$. Under these assumptions u_i/σ_u is standard normal and by the symmetry of the normal distribution one can write the probability of Year 12 completion as:

$$P(y = 1|\,\mathbf{Z}) = \Phi\left(\mathbf{X}\frac{\boldsymbol{\beta}}{\sigma_u} + \frac{\gamma}{\sigma_u}LC^* + \frac{\boldsymbol{\theta}}{\sigma_u}\mathbf{W}\right), \tag{2}$$

where $\Phi(\cdot)$ is the cumulative distribution function for a standard normal. Equation 2 describes the well-known probit model.[19] In the models considered in this paper, however, LC^* is a latent continuous variable representing youth's locus of control. LC^* is assumed to be distributed $N(0, \sigma_\ell^2)$. The higher the value of LC^*, the more internal an individual is and the lower the value, the more external an individual is.

Although the econometrician does not observe LC^*, imperfect measures of LC^* are observed, called say, l_j for $j = 1, 2, ..., J$. In reality the l_j are not observed: what is observed are ordered responses. The latent locus of control measure LC^* relates to these indicators through the following *measurement model*:

$$l_j^* \;=\; \alpha_j LC^* + \epsilon_j \;\;;\;\; j = 1, 2, ..., J. \tag{3}$$

In this set of equations α_js are parameters to be estimated and ϵ_j represent j error terms such that conditional on LC^*, $\epsilon_j \sim N(0, 1)\;\forall j$. The error terms in this system of J ordered probits are independent of each other and from the error term in the Year 12 completion equation (e.g. $E[\epsilon_j \epsilon_i] = 0\;\forall j \neq i$ and $E[\epsilon_i u] = 0$).[20] The rule linking the ordered responses

[19] Note that in this model only the ratio β_i/σ_u is identified. The standard normalisation is $\sigma_u = 1$.

[20] Note that by assuming $\epsilon_j \sim N(0, 1)$, the model uses a common identification restriction; that is $\sigma_{\varepsilon_j}^2 = 1$.

to the latent l_j^* is:

$$l_j = \begin{cases} 0 & \text{if} \quad -\infty < l_j^* \leq \delta_{1j}, \\ 1 & \text{if} \quad \delta_{1j} < l_j^* \leq \delta_{2j}, \\ \vdots & \qquad \vdots \\ M_j & \text{if} \quad \delta_{Mj} < l_j^* < \infty; \end{cases} \tag{4}$$

where δ_{ij} $\forall i = 1, 2, ..., M_j$ and $\forall j = 1, ..., J$ are threshold parameters satisfying the restriction $\delta_{1j} < \delta_{2j} < ... < \delta_{Mj}$ $\forall j = 1, ..., J$. In this setup $M_j + 1$ denotes the total number of categories, or possible answers, for indicator j. In the survey there are four possible answers for each indicator (ranging from "*Strongly Agree*" to "*Strongly Disagree*") which implies that $M_j = 3$ $\forall j$. There are seven indicators of locus of control in the survey, $J = 7$.

The objective is to obtain estimates of $\boldsymbol{\beta}$, α_js, thresholds for each of the ordered probit models (δ_{ij}s), and the main parameters of interest, γ and $\boldsymbol{\theta}$. In addition to the standard restrictions on the variance of the error terms made in the probit and the system of ordered probits detailed above, to identify the parameters in the model it is also necessary to set one of the α parameters to unity.[21] Note that (abstracting from the discrete nature of the indicators) an alternative interpretation of the model is similar to the case where a measurement error problem in an explanatory variable is solved by including one of the indicators in the regression and using the others to instrument it. The analogy is nevertheless not completely accurate in this case because of the non-linearity of the probit model.

Note that the model can be seen as a system of J ordered probits (given by the set of equations in Equations 3 and 4) and a binary probit model for Year 12 completion (Equation 2). The system has cross-equation restrictions on some of the parameters and a common factor with known distribution (LC^*). The parameters in this model are estimated by maximum likelihood with adaptative quadrature for the numeric maximization of the likelihood. This is done using the software aML (Lillard and Panis, 2003), but any other software that performs maximum likelihood can also be used.[22]

[21] This is a standard normalisation in the literature. An alternative normalisation is to set the variance of LC^* to unity.

[22] aMl software is freely available from http://www.applied-ml.com.

In summary, with the above model it is possible to test, on one hand, whether there is a relationship between locus of control and Year 12 completion while taking into account the latent nature of locus of control, and on the other, whether growing up in a socio-economically disadvantaged household is related to the probability of completing Year 12. Based upon the literature review in section 2, it is expected that $\hat{\gamma} > 0$ and $\hat{\theta}_i < 0$ for all $i \in \theta$.

4.2 A probit model with selection and latent locus of control

The second educational outcome is whether the youth obtained an ENTER score. To obtain an ENTER score students have to complete a pre-university entrance program during their last year of education (Year 12). The ENTER scores are based on state-wide examinations and on results for specific subjects taken during Year 12.[23] The score obtained is the percentile rank of a student's performance within their own cohort.

The econometric model, however, is not as straightforward as the previous model for Year 12 completion because the researcher is not able to observe the decision to obtain an ENTER scores for youths who did not complete Year 12. Ideally, the researcher would like to estimate an "obtained ENTER score" equation for all school youths. This equation would represent all school students in Year 12, whether or not these students had finished Year 12 at the time of the interview. But there is a sample selection problem as the researcher only observes whether those students who had completed Year 12 at the time of the interview, had obtained an ENTER score. That is, there might be a non-random sample selection because of incidental truncation (e.g. whether people would obtain an ENTER score is missing as a result of Year 12 non-completion). Because the propensity to complete Year 12 may be correlated with unobservables that affect the propensity to obtain an ENTER score (ability being a prime example) using only students who completed Year 12 could produce inconsistent estimates.

To take into account the non-randomness of the sample used in estimating the

[23] These requirements vary across states in Australia. In some states, for example, there is no state-wide examination, and only results from specific subjects taken in Year 11 or 12 are considered. See (Marks et al., 2001, Appendix 3) for a detailed description. There is no unified name across Australian States for what is called ENTER score in this paper. In addition, the scale for ENTER scores for Queensland takes values from 1 to 25, where 1 represents the highest ranked students. The ENTER score used in the present analysis transforms Queensland scores to scores equivalent to other states. The ENTER scores, or entrance ranks, are calibrated to a common, Australia-wide scale that ranges from 30 to 99.99 (see Cardak and Ryan, 2006).

"obtained ENTER score" equation, the model considered is a probit model of whether the youth obtained an ENTER score with a probit selection equation for Year 12 completion. Explicitly, the model is as follows:

$$ENTER = 1\left[\mathbf{X_1}\boldsymbol{\beta_1} + \gamma_1 LC^* + \boldsymbol{\theta_1}\mathbf{W} + u_1 > 0\right] \tag{5}$$

$$HS = 1\left[\mathbf{X_2}\boldsymbol{\beta_2} + \gamma_2 LC^* + \boldsymbol{\theta_2}\mathbf{W} + u_2 > 0\right] \tag{6}$$

where $1\left[\cdot\right]$ is the indicator function, the second equation is the sample selection equation and $ENTER$ is only relevant when $HS = 1$. Here, (u_1, u_2) is independent of all explanatory variables, distributed bivariate normal with zero mean, and unit variances.[24] The correlation between u_1 and u_2 is denoted by ρ. Note that $\mathbf{X_1}$ is a subset of $\mathbf{X_2}$ due to the exclusion restriction. Since in this model LC^* is a latent variable, the model is complemented with a set of equations as in Equations 3 and 4. The error terms in the measurement equations in Equation 3 are assumed independent of each other and independent of u_1 and u_2. Estimation of this model is by maximum likelihood.

4.2.1 The exclusion restriction: school starting age rules

What remains to complete the description of the above model is to define the variable (or variables) that constitute the exclusion restriction. Although the model above is identified without this restriction, identification is off of the non-linearities in the probit models, and hence not very convincing. A more convincing analysis involves at least one variable that determines selection (i.e. affects Year 12 completion) but does not partially affect the likelihood of obtaining an ENTER score. The exclusion restriction in the present analysis is based on a combination of school starting age rules and years of education required in Australian states.

In Australia, each state has the power to establish rules about school starting age in their territory. Although in general all children start school when they are five years old (e.g. either kindergarten or Year 1), at the beginning of the school year some children are four years old (and then turn 5 during the school year) while some others are 5 years

[24] The selection equation is exactly the same as the Year 12 completion model described in the previous section.

old (and turn 6 during the school year). In addition, some states include kindergarten as the normal progression of the school experience. States requiring kindergarten have 13 years of education in total—one year more than those states that do not. Table 4 presents school starting age rules affecting the members of the YIF cohort. The table also presents information on the number of years of education required in each state. Under normal progression, the YIF cohort started school (kindergarten) in 1993, except for those who lived in Queensland, Northern Territory, and Western Australia. In these three states kindergarten is not required, and youths in the survey would have started Year 1 in either 1994 or 1995—one and two years later than in the other states. This is the case because the school starting age rule for these states falls right in the middle of the time at birth of the YIF cohort (October 1987 to March 1988). The combination of the starting age rules and the number of years of education implies, according to the last column in Table 4, that youths who were born in Queensland, the Northern Territory, and Western Australia in the second three months of the cohort period (January through March 1988) would have graduated a year later than those youths from the same states (and any other state) who were born between October and December 1987.[25]

[Table 4 here]

Table 5 presents the unconditional probability of completing Year 12 by (i) state and (ii) whether youth were born in the first or the second part of the cohort period. The difference in probabilities and p-values for a test of equality of means are presented in Column 3. The table is also organised according to whether or not state rules in school starting age were in place for the cohort. In states in which the rules coincided with the YIF cohort (Queensland and Western Australia: call these Policy States) those youth who were born in the first part of the cohort period (Oct. to Dec. 1987) are, as expected, more likely than those born later in the same states (Jan. to Mar. 1988) to have completed Year 12 by 8.2 and 7.6 percentage points. These differences are statistically significant at the 10 percent level.

[25] Although the difference at the start of school were one and two years, the difference at the end of Year 12 is only one year because in these three states there is one less year of education. Because of the small number of observations, the Northern Territory sample was added to the Western Australia sample, even though the rules for the two do not coincide exactly.

[Table 5 here]

Surprisingly, in states where rules did not coincide with the YIF cohort period (and hence no differential effect is expected on Year 12 completion rates) Table 5 reveals a higher, and in several states statistically significant, probability of Year 12 completion when being born early in the cohort (Oct. to Dec. 1987). Although results for the ACT and Tasmania are based on small samples (57 and 135 respectively) and hence should be interpreted with caution, results for the other states reveal a higher probability of graduation for those who were born early in the sample. Being born early (Oct. to Dec. 1987 as opposed to Jan. to Mar. 1988) in New South Wales and Victoria is associated with a 7.2 and 16.1 percentage points higher probabilities of completing Year 12.[26] Unexpectedly, results for these states are consistent with scenario in which parents initiate children's education experience as soon as they can and the school system allows them to slightly stretch the rules.

Consequently, the exclusion restriction in the main probit equation is a interaction between living in Queensland, Western Australia, or the Northern Territory and a dummy variable indicating whether the youth was born between October 1987 and January 1988. This interaction reflects the state rules about school starting age discussed above.[27] There is no obvious reason to think that this variable affects the decision to obtain an ENTER score. This exclusion restriction meets the two formal requirements for a good instrument: (i) it is correlated with Year 12 completion (as explained in this section), and (ii) there is no compelling reason to believe that it belongs in the equation for obtaining an ENTER score. Given these, the exclusion restriction contributes in dealing with the selection problem in the present case because the proposed instruments are related to the selection equation, but not to the main equation.

[26] To verify that these results are not due to outliers I recalculate these probabilities changing the cut-off date. The cut-off date is varied from 15 October 1987 to 15 March 1988 and the difference in probability is calculated at each day in this period for each State. Appendix Figure A2 shows the results.

[27] A model in which the exclusion restriction is a dummy variable of whether the youth were born in the first part of the cohort period (Oct. to Dec. 1987), and including state dummy variables was also estimated. The coefficient on this dummy variable is highly significant at the 5 percent level.

4.3 A censored regression model for ENTER scores

The third and last educational outcome considered in this paper is the ENTER score reported by youth. The study and interpretation of this model is conditional on getting an ENTER score. Because the actual ENTER score is left-censored at 30 (e.g. youth who obtained an ENTER score below 30 are reported as having obtained 30), the model used is a censored (normal) regression model.[28] Unlike running an OLS regression of the observed ENTER scores on the regressors of interest, the censored regression model yields consistent estimators of the parameters. As with the models for the other two educational outcomes, the censored model treats locus of control as a latent variable by having attached to it a measurement model as in Equations 3 and 4. The censored model is:

$$ENTER^* = \mathbf{X}\boldsymbol{\beta} + \gamma LC^* + \boldsymbol{\theta}\mathbf{W} + u \tag{7}$$

$$ENTER = max\,(30, ENTER^*) \tag{8}$$

where $ENTER^*$ is the actual ENTER score, but it is only observed if it is greater than the censored value of 30. In this specification u, conditional on regressors and censored value, is assumed to be normally distributed with mean zero and variance σ_u^2. Regressors are as described in previous specifications. An additional difference between this model and the Tobit model is that the estimates are directly interpretable. Estimation is by maximum likelihood.[29]

5 Results

This section presents the estimation results for the models described in the previous section. Before discussing the main results, it is necessary to establish if the measurement

[28] Notice that this model is similar to a Tobit model, but the Tobit model reflects economic behaviour that produces zeros (or other values). In the censored regression model, the underlying variable is also continuous but it is censored due to data collection or institutional arrangements.

[29] The original specification of this model included a selection equation as in the model for obtaining the score. The nature of the selection in this case is, as before, incidental truncation; the researcher only observes the ENTER score for those who decided to take it. It was attempted to estimate the model with selection but it proved very difficult to make the maximisation algorithm to converge. If, as most likely, people who do not take the exam are of lower ability than those left in the sample (and as believed in the economics literature) ability is correlated with locus of control, one would expect the estimates presented in the results section to be an under-estimate of the true parameters. That is because there are fewer people with an external locus of control in the sample.

part of all models provides estimates that are consistent with the interpretation of internal versus external locus of control. Recall that α-parameters in Equation 3 link the latent locus of control and the observed (categorical) indicators. For the locus of control interpretation assumed in the previous section to be consistent with the data (e.g. higher values represent a more internal locus of control), the set of parameters α need to be positive for those statements suggesting an internal perception of control, and negative for those statements suggesting a more external perception of control.[30] Finding significant estimates would indicate a strong association between the latent variable (locus of control) and the observed set of indicators.

Table 6 reports maximum likelihood estimates of these parameters (α) for each of the locus of control questions and for the three educational outcomes under consideration. Although each model is estimated jointly, this table presents only the measurement part of each of them. Figures in parentheses are heteroscedasticity-robust standard errors. Table 6 also reports the standard deviation estimate of the latent locus of control variable. Since the latent locus of control has no intrinsic units of measure it is necessary to set one of the α parameters to a constant. In this case, the coefficient on the first indicator is set to unity (i.e. I cannot solve some of my problems).[31] These estimates reveal how strongly associated the latent locus of control is to each of the observed indicators. The table shows that all parameters are statistically significant and have the appropriate sign for an interpretation in which higher values correspond to internals and lower values to externals. That is, a youth who strongly disagrees with the statement that they feel pushed around in life (i.e. the indicator takes the highest value, 4) will have higher values of locus of control because the coefficient relating the two is positive and significant. The same holds for all statements in the question where disagreeing is a sign of internality (lines 2, 3 5, and 7). For the other two statements in which disagreeing is a sign of an external locus of control (line 4 and 6 in the table), the coefficients are negative; this implies that all coefficients are consistent with the same interpretation once the phrasing of the statement is taking into account (i.e. low values of LC^* for external and high values for internals). Additionally, parameters in all three models are very close to each other (this

[30] The locus of control variables are coded as: 1 Strongly Agree, 2 Agree, 3 Disagree, and 4 Strongly Disagree.

[31] Another alternative is to set the variance of LC^* to unity.

is unsurprising since the measurement part of the models is similar for all outcomes). Finally, the estimate of the standard deviation of the locus of control variable in the last row of the table is necessary to calculate marginal effects due to changes in locus of control. In particular, since the model assumes that the latent locus of control is distributed normal with mean zero and variance σ_ℓ^2, all that is needed to obtain values at different percentiles of the locus of control scale is an estimate of the variance. With this, the researcher is able to ask how much the probability of completing Year 12 would change should a person becomes more internal, for example, by moving from the 25$^{\text{th}}$ percentile of the locus of control distribution to the 75$^{\text{th}}$ percentile.

[Table 6 here]

5.1 High-school completion

Table 7 presents marginal effects of changes in explanatory variables for all three educational outcomes. Statistical significance, denoted by stars, is based upon the significance of the underlying parameter in the probit models (see Appendix Table A3 for raw parameter estimates).

[Table 7 here]

The estimates in Table 7, column 1, suggest that becoming more internal, as measured by moving from the 25$^{\text{th}}$ to the 75$^{\text{th}}$ percentile of the locus of control scale, increases the probability of completing Year 12 by 6.1 per cent.[32] This estimate is only statistically significant at the 10$^{\text{th}}$ per cent level, but economically sizeable. It is of almost the same size as for growing up in a household where both parents live together (e.g. the effect in the latter case is 6.5 per cent).

The stratification in the YIF data and the modification introduced in this paper (i.e. combining Stratum D and F), allows two types of comparisons. First, it is possible to compare people extremely disadvantaged—those categorised as Stratum B, six years or more of parental income support— to (a) non-disadvantaged youths (Stratum A) and (b) those intermediately disadvantaged (those disadvantaged for less than six years, Stratum

[32] Becoming more internal by one standard deviation ($\hat{\sigma}_l = 1.031$) increases the probability of completing Year 12 by 2.2 per cent.

C, D and F, and E). The second comparison allows a partial look at the timing of exposure to disadvantage and its effects on youth's educational outcomes by looking at the difference between Strata C, D and F, and E. It can only be a partial look because these three variables are defined conditional on being exposed to disadvantage at home for less than six years. Given the stratification variable, it is not possible to look at the timing issue for families that received income support for more than six years (that would entail splitting the Stratum B). The results for intensity follow below, while the timing issue is investigated in the Extensions section.

When compared with youth from non-disadvantaged families (Stratum A), youths growing up in socio-economically disadvantaged households have lower Year 12 completion probabilities (all other Strata). Youths exposed to any degree of disadvantage are between 3 and 10 per cent less likely to complete Year 12 than youths from non-disadvantaged families. Youth exposed to disadvantage for more than 6 years exhibit the lowest probability of finishing school, 9.9 per cent. Conditional on being only intermediately exposed to disadvantage (Strata C, D and F, and E), the figures also seem to support the hypothesis that the age at which children are exposed to disadvantage doe matter: relative to non-disadvantaged youth, those who were exposed when 6 to 10 years old show the lowest probability, 8 per cent, of completing Year 12. Exposure to disadvantage when less than 6 years old and when older than 10 years results in economically sizeable but statistically insignificant effects on the chances of completing Year 12. This seems to imply that being intermediately disadvantaged doe matter, but being disadvantaged when 6 to 10 years old matters more. The Extensions section investigates these differences formally.

In addition to the strong association between parental disadvantage and Year 12 completion, individual and family characteristics variables have sizeable and statistically significant effects on the probability of graduation. Consistent with the literature, parental education is associated with greater school graduation probabilities. Surprisingly, there is no effect of parental non-monetary investments on their children's propensity to complete Year 12 (as measured by whether parents read to them at night and whether parents helped them with homework). Indigenous youth are also 17 percentage points less likely to complete Year 12; whilst males are 12 percentage less likely to

graduate than females. Youths with at least one parent who is an immigrant from a non-English-speaking country exhibit a higher propensity to complete Year 12 by almost 12 percentage points; whilst youths with either parents an immigrant of English-speaking background show no significantly different probability of graduating than youths whose parents were born in Australian. These figures are in line with previous research that finds that the children of non-English speaking background immigrants tend to have higher graduation rates than either Australian-born or English-background immigrants (see, for example, Larum and Beggs, 1989).

Finally, the model for Year 12 completion includes an interaction term between (i) being born in Queensland or Western Australia and (ii) whether youths were born between October and December 1987 (early born). As discussed in the description of the model, this constitutes the exclusion restriction that used to face the selection problem due to incidental truncation for the outcome of obtaining an ENTER score. Youths who were born in the period October to December 1987 in Queensland and Western Australia are 3.6 percentage points more likely to graduate than youths who were born in the same states but in the period January to March 1988. This effect is expected, due to the rules on school starting age in these states. Surprisingly, youths who were born in the period October to December 1987 in all other states are 16.7 percentage points more likely to complete Year 12 than youths born later (January to March 1988) in those states. This result is puzzling given existing school starting age rules: one would expect a bigger difference in states where these rules affected differentially some members of the cohort (QLD and WA). The results indicate, nevertheless, that being born just a few weeks later might affect substantially the probability of Year 12 completion in all states. A likelihood ratio test rejects the hypothesis of the insignificance of the coefficients on the interaction term and the Early Born dummy at any standard level of statistical significance.[33]

[33] A model that included only a dummy variable for being born early (Oct. to Dec. 1987) and state dummy variables was also estimated. The effect of being born in the first three months of the cohort period (Oct. to Dec. 1987) increases the probability of completing Year 12 by almost 10 per cent. This coefficient is statically different from zero at the 5 percent level. This also renders support for the hypothesis that being born early in the sample has a substantial impact on the probability of completing Year 12.

5.2 Obtaining an ENTER score

This section discusses estimation results of the model for obtaining an ENTER score. Because it is a concern that the sample of people who finish Year 12 (and hence are able to apply for an ENTER score) is non-random, the analysis uses the model for Year 12 completion, and the exclusion restriction previously discussed, as the selection equation. The second column of Table 7 reports changes in the probability of obtaining an ENTER score given marginal or discrete changes in the explanatory variables (see Column 2 in Appendix Table A3 for probit models estimates).

The table reports a correlation between the error terms in the latent specification of the probit models as 0.551 with standard error 0.240. A formal test of the significance of this parameter rejects the null hypothesis that it is statistically equal to zero at the 5 per cent level. This suggests that selection is present and needs to be taken into account. The figures in Table 7 also suggest that becoming more internal (moving from the 25^{th} to the 75^{th} percentile of the locus of control distribution) increases the probability of obtaining an ENTER score by 7.6 percentage points conditional on completion. Youths who are more internal tend to have a higher inclination to obtain an ENTER score and potentially attend university.

Parental socio-economic status, and being disadvantaged at home, has a big impact on the decision to obtain an ENTER score. The most disadvantaged youth (six years or more of parental income support use) are 19.8 percentage points less likely to obtain a score. For those disadvantaged for less than six years, results indicate that exposure to disadvantage at younger ages has a significant effect on the choice of obtaining an ENTER score (10 per cent for those aged 6 to 10 , and 11.6 for those aged less than 6). Youth who were exposed to disadvantage when they were 10 to 16 years old are 3.7 per cent less likely to obtain a score, although this is not statistically significant.

As shown in Table 7, parental and youth characteristics show a sizeable association with obtaining an ENTER score. There is evidence, for example, that male and indigenous youth are 6.6 and 22.5 percentage points respectively less likely to obtain a score than are their counterparts. Surprisingly, youth who lived with both parents at age 15 show no higher probability of obtaining a score. Youths whose parents read to them

before going to bed seem to develop educationally-consistent skills that might allow them to build higher expectations in terms of tertiary education (they are 6.4 percentage points more likely to obtain an ENTER score). The effect of parents helping youth with their homework, however, is unexpectedly negative and significant (6.3 points), canceling almost exactly the benefits of reading at night. The negative coefficient on this variable, however, is consistent with the interpretation that low ability youths might be more likely to get help from their parents. Finally, parental education, for both mother and father, shows a strong association with youths decision to obtain an ENTER score, as does being a youth whose parents immigrated to Australia from a non-English speaking country. Qualitatively, most of the results are similar to the results for Year 12 completion.

5.3 Results for ENTER scores

This section discusses the results of a censored regression model for ENTER scores. The use of a censored model is necessary because for youths who obtained an ENTER score lower than 30 the authorities deemed the score to be equal to 30. There are 17 youths in the sample with an ENTER score equal to 30 (and 4 with the maximum value 99.99). As discussed previously, this model also has attached a measurement model for the locus of control variable.

These results are conditional on youths having completed Year 12 and having obtained an ENTER score. Unfortunately, a model which tried to take into account the non-randomness of the data, and hence tackle potential selection problems due to incidental truncation, was tried but the maximisation algorithm did not converge.[34] If, as most likely, youths who did not obtain an ENTER score are of lower ability than those remaining in the sample, and as believed in the economics literature, ability is correlated with locus of control, the estimates that follow will under-estimate the relationship between locus of control and performance in the population as a whole. Results, hence, are interpreted as being conditional upon Year 12 completion and obtaining an ENTER score.

Is locus of control associated with performance? To briefly explore this issue col-

[34] Several strategies to determine initial values were tried. It might have just been the case that the model was too complex. It was a Censored Regression as the main equation with a probit selection equation for Year 12 completion, in addition to the measurement part.

umn 3 in Table 7 presents regression estimates of the relationship between ENTER scores and the set of explanatory variables.

The results indicate that becoming more internal by one standard deviation (σ_ℓ) increases ENTER scores by 0.94 points (column 3). This result is small and statistically significant at the 10 per cent level. Additionally, youths who grew up in the most socio-economically disadvantaged households obtained, on average, ENTER scores that are 4.26 percentage points lower than youths from non-disadvantaged households. Youths with intermediate exposure to disadvantage at home, however, show no statistically lower ENTER scores (Strata C, D and F, E) conditional on having obtained an ENTER score.

Parental characteristics seem particularly relevant for performance. Parental involvement in youths' education (either by reading to them at night or by helping them with homework) shows a strong association with better performance (as reflected by a higher ENTER score). Parental education also correlates positively with better performance, as does having immigrant parents from a non-English speaking background.

6 Extensions

What is the interaction between disadvantage and locus of control?

In the simple cross-tabulations presented when introducing the data, it was shown that locus of control and disadvantage were in fact correlated. This section investigates whether there is evidence of a potential indirect effect of parental disadvantage on the educational outcomes of youth. This section presents the results from a model in which (latent) locus of control is regressed on the disadvantage dummy variables, both with and without other controls. The coefficients for the disadvantage variables in that model are presented in Appendix Table A5. The models in both columns take into account the latent nature of locus of control and use the variables in the survey as imperfect measures of it. The model is similar to the one estimated for the educational outcomes, but in this case locus of control is the dependent variable. With no controls, the results suggests that being highly disadvantaged (Stratum B) is associated with being more external (as lower values of the locus of control latent variable indicate externality). This relationship is statistically sig-

nificant at the 5 per cent level. Being intermediately exposed to disadvantage (Stratum C, D and F, and E), however, shows no statistically significantly relation with locus of control. When introducing controls into this regression, such as parental education, demographic characteristics of youth, and parental immigration status among others, the coefficient reduces (in absolute terms) from -0.150 to -0.116 and becomes statistically insignificant. Given these results it is very unlikely that there is an indirect effect of disadvantage on youths' educational outcomes through the effect of disadvantage on locus of control, basically because the latter is very small.[35] This means that there is no evidence in support of theories that suggest that disadvantage "produces" individuals with different personality characteristics than those growing up in non-disadvantaged households once other factors, such as parental characteristics, have been taking into account.

Disadvantage and educational outcomes: a closer look

The results from the models presented in the previous sections suggest a negative relationship between educational outcomes and parental socio-economic background (i.e. disadvantage). This section tests three hypotheses of interest. First, it tests whether youths growing up in non-disadvantaged households and those subject to any disadvantage at home differ in their educational outcomes. This is carried out by testing the hypothesis that all coefficients on the disadvantage dummy variables are jointly equal to zero. The second hypothesis is whether highly disadvantaged youths (Stratum B) show similar educational outcomes as intermediately disadvantaged youths (those in Strata C, D and F, E), all relative to the base group (i.e. the non-disadvantaged group). In practice, the test is whether the coefficient on the highly disadvantaged dummy (θ_B) individually equals each of the intermediate disadvantaged categories. That is, Stratum B versus (i) C, (ii) D and F, and (iii) E. Finally, this section tests the hypothesis that the effect of disadvantage on youths' educational outcomes varies with the age at which youths were exposed to disadvantage at home. As previously mentioned, due to the design of the survey it is only possible to test this hypothesis for periods of disadvantage that last for less

[35] Models for educational outcomes with interaction terms between locus of control and each of the parental disadvantage dummy variables were estimated. The coefficients on the interaction terms in all three models (not shown) were all small and statistically insignificant at the 10 per cent level. This evidence, therefore, does not lend support to the hypothesis that parental disadvantage has an indirect effect on the educational outcomes of youth through its effect on youths' personality.

than six years. To implement this, the test is whether the coefficients on the intermediate disadvantage dummy variables differ from each other.

Appendix Table A4 reports Likelihood Ratio tests (LR-statistics) and their corresponding p-values for the hypotheses just described.[36] The first row in the table presents the results of a joint test for the hypothesis that there is no effect of parental disadvantage on the three educational outcomes considered (all relative to the base group—the non-disadvantaged group). As expected, the test rejects the hypothesis at the one per cent level of significance. Therefore, being disadvantaged has an effect on all educational outcomes. From the negative coefficient in the original specification it is possible to assert that the effect is negative. This is not surprising since the results from the baseline model indicated that being *highly* disadvantaged was strongly associated with the worst educational outcomes. The second panel in the table (rows 2, 3, and 4) presents results for the hypothesis that the effect of being highly disadvantaged is similar to the effect of being intermediately disadvantaged. The results suggest that being highly disadvantaged (Stratum B) is as damaging as being disadvantaged for a shorter period and at different periods in the youth's life (all compared to the non-disadvantaged group). I find that the effect of being highly disadvantaged (Stratum B) is statistically different from being disadvantaged for less than six years and when between 10 and 16 years old (Stratum C). This result holds for all three outcomes. Additionally, the data also reject the hypothesis that being highly disadvantaged is different from being intermediately disadvantaged (Stratum D and F, and Stratum E) in the model for obtaining an ENTER score and in the censored model for ENTER scores. In the Year 12 completion model there is no evidence to reject these hypotheses.

Finally, the last three rows in the table report test results for the hypothesis that the effect of disadvantage on educational outcomes varies with the age at which youths are exposed to disadvantage at home (at least for those exposed to disadvantage for less than six years). Results show that there is no evidence to indicate that the effect of disadvantage on educational outcomes varies by the age at which youths experience disadvantaged at home. The statistical tests cannot reject the hypotheses that the effect is the

[36]The models used are the main models discussed in previous sections. Restricted models for these tests of hypotheses are not shown.

same in each of these tests.

Overall, these tests present evidence to support the hypothesis that disadvantaged youths exhibit worse educational outcomes that non-disadvantaged youths. It is more difficult however to disentangle the relative effects of being highly and intermediately disadvantaged when young. There is no evidence in support of the hypothesis that the age at which youths are exposed to disadvantage alters the relationship between disadvantage and the youths educational outcomes.

How does the measurement model compare with traditional approaches of dealing with latent variables for which several (imperfect) measures are available?

In this section tries to answers the following question: how much do results change if instead of using the measurement model one were to use the standard approach in the literature and in some way combine the locus of control variables into a single index? Following the standard approach (i.e. adding the seven different values for each individual and then standardising the resulting variable to have mean zero and variance one) yields a probit estimate of 0.062 (SE=0.033) for the locus of control variable in the Year 12 completion model. Comparing this estimate with the estimate from a measurement model in which the variance has been set to one reveals only a small difference (Coeff.=0.059; SE=0.036). The use of an arbitrary index in this case seems to slightly inflate the correlation between locus of control and the probability of completing Year 12.[37]

The small difference between these two estimates is puzzling as it was expected to be larger. One explanation for this might be that the arbitrary index looks normally distributed, and in the measurement model it is assumed that the latent locus of control is normally distributed. It is unknown, however, how the results from comparing these two types of approaches might change when relaxing some of the assumptions in the measurement model. The question then is: if using the measurement component in the models for educational outcomes does not provide large differences in estimates, why should one use it at all? First of all, it was unknown beforehand that the results were similar, and second, the measurement model makes explicit assumptions about the nature of the concept and its measurement, and also provides estimates of the relationship

[37] Similar results were found for the other two outcomes.

between the latent variable and its observed (imperfect) indicators. These types of comparisons are interesting given that it is common to find these type of arbitrary indexes in the literature, although the effect of the assumptions behind these alternatives is unknown.

7 Conclusions and final remarks

This paper focuses on investigating the factors associated with three educational outcomes for a cohort of Australian Youth: (i) whether they complete school and (ii) whether they obtained an ENTER score, and (iii) their actual ENTER score. The paper focuses on two of these factors: locus of control and parental socio-economic background.

The results suggest that individuals with a more internal locus of control are more likely to complete Year 12, obtain an ENTER score, and perform better at school (as measured by their actual ENTER score) than youths with a more external locus of control. The methodology employed took into account the latent nature of the locus of control concept and uses survey data as imperfect measures of it. Some caveats, however, apply to these results. First, the timing at which the questions on locus of control are asked is not the best for the purposes of this paper. Ideally, one would like to relate educational outcomes to a measure of locus of control taken earlier in life, or at least several years before the outcomes. The concern is that educational outcomes might affect individuals' locus of control.[38] Once the second wave of the YIF data becomes available in early 2009, it will be possible to evaluate other outcomes measured several years after the initial measure of locus of control (e.g. university enrolment, living arrangements, further studies). It will also be possible to compare how the locus of control has changed since the first interview. Earlier results in the literature, particularly from the United States, indicate that becoming more internal increases the probability of Year 12 completion (by one standard deviation) of between 2 and 3 per cent (Coleman and DeLeire, 2003). The results in this paper are similar to those estimates. This suggests that the results might not be as biased as expected.

[38] This is not an issue if, as psychologists believe, locus of control is fully developed by the time an individual is 18 years old. Although economists are sceptical of this interpretation, results from the happiness literature show that shocks to happiness have only temporary effects.

A second caveat for the results involving locus of control involves the absence of ability in the regressions reported in this paper. In the economics psychology literature researchers are concerned with the possibility that locus of control might be positively correlated with ability. If ability is not controlled for and locus of control is indeed correlated with ability, it is likely that locus of control is endogenous to the outcomes under study. Apart from the ENTER score, there is no other measure of ability available in the Youth Questionnaire in the YIF data. The ENTER however is only available for those who had finished school by the time of the interview. Given this expected correlation, the results presented in this paper might over-estimate the relationship between locus of control and educational outcomes.

This paper also analyses the relationship between parental socio-economic background and youths' educational outcomes. The contribution of the paper in this regard is to provide an alternative estimate based not on parental occupation at some point in time, but on almost 12 years of parental income support histories. The results indicate that even after controlling for individual and other parental characteristics, youths who grew up in highly disadvantaged households (e.g. parents' income support use lasted for more than six years) are almost 10 and 20 percentage points less likely to complete Year 12 and obtained an ENTER score, and obtained ENTER scores 4 points lower than non-disadvantaged youth. As education outcomes are valued and rewarded in the labour market, these youth would most certainly be disadvantaged when entering it. Although it seems sensible to assume that parental socio-economic status is exogenous to the educational outcomes of youth, unobserved family-specific heterogeneity might confound interpretation of the results. Nevertheless, by using data on siblings to account for unobserved family characteristics, Ku and Plotnick (2003) suggest that cross-sections estimates of the relationship between disadvantage at home and school completion of youth are biased downwards. If this is the case in the Australian context, the results discussed above under-estimate the effect of disadvantage on the educational outcomes of youths.

This paper also investigates the relationship between locus of control and disadvantage. The concern is that disadvantage might potentially affect the personality characteristics of youth and in that way indirectly affect educational outcomes. By running a regression of locus of control on the disadvantage measures, this paper finds that only

highly disadvantaged youths have a more external locus of control. The small difference between highly and non-disadvantaged youth, however, disappears after controlling for other family characteristics, making unlikely the existence of an indirect effect of disadvantage on educational outcomes through its effects on locus of control.

Given the uniqueness of the data, this paper also explores the hypothesis that youth who are exposed to disadvantage at different times of their lives might be affected in different ways in terms of their educational outcomes later in life. There is no evidence to support this hypothesis.

The present paper contributes to the literature in several ways. First, it provides some evidence of the relationship between locus of control and educational outcomes in the Australian context. This provides some evidence on the relationship between personality and educational outcomes. The models used in this paper incorporated the view that locus of control is a latent variable and that the researcher only observes imperfect measures of it. Second, the analysis provides alternative measures of the relationship between disadvantage and parental socio-economic status (SES). Traditional measures of SES in Australia are based on parental occupation at some point in the life of the youth. This paper uses as a measure of parental background a classification based on almost 12 years of parental income support history. Additionally, the paper explores the hypothesis that disadvantage affects the personality characteristics of children (e.g. locus of control), and in that way indirectly affects educational outcomes (locus of control is traditionally used as a regressor and not as a dependent variable). Finally, it was found that although rules about school starting age in Australia are in place and apply only to some members of the YIF cohort, there is a high (almost 10 percentage points) and statistically significant probability that youth who were born at the end of the year would graduate earlier than youth who were born in the first months of the following year. This clearly deserves further investigation as the results suggest that this is true not only in states where the policy leads us to expect this, but all across Australia.

All in all, the results in this paper suggest the existence of a positive association between being more internal and positive educational outcomes for Australian youth. It also shows that highly disadvantaged youths are alarmingly more likely to not finish school (or finish it later) than non-disadvantaged youth, and those who finish perform

relatively worse in the ENTER score. The difference between the educational outcomes of non-disadvantaged and disadvantaged youth will more likely accentuate the disparity in labour market outcomes of these two groups during the course of their lives.

References

Altonji, J. G., T. E. Elder, and C. R. Taber (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *The Journal of Political Economy 113*(1), 151–184.

Andrisani, P. J. (1977). Internal-external attitudes, personal initiative, and the labor market experience of black and white men. *The Journal of Human Resources 12*(3), 308–328.

Andrisani, P. J. and G. Nestel (1976). Internal-external control as contributor to and outcome of work experience. *Journal of Applied Psychology 61*(2), 156–165.

Becker, G. (1993). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. Chicago, IL, USA: The University of Chicago Press.

Bowles, S., H. Gintis, and M. Osborne (2001). The determinants of earnings: A behavioral approach. *Journal of Economic Literature 39*(4), 1137–1176.

Breunig, R., D. Cobb-Clark, T. Gorgens, and A. Sartbayeva (2007, September). User's guide to the Youth in Focus Project, version 1.0. Youth in Focus Project Discussion Paper Series No. 1, The Australian National University, Research School of Social Sciences (Economics Program), Canberra, Australia.

Cardak, B. A. and C. Ryan (2006, June). Why are high ability individuals from poor backgrounds under-represented at university? Discussion Paper No. A06.04, La Trobe University, Melbourne, Australia.

Carneiro, P., C. Crawford, and A. Goodman (2007, October). The impact of early cognitive and non-cognitive skills on later outcomes. CEE Discussion Papers 0092, Centre for the Economics of Education (CEE), LSE, London, UK.

Cebi, M. (2007). Locus of control and human capital investment revisited. *Journal of Human Resources 42*(4), 919932.

Centrelink (2007). A guide to Australian government payments. `<http://www.centrelink.gov.au/internet/internet.nsf/filestores/>` (accessed January 2008).

Clarke, J. H., B. V. MacPherson, and D. R. Holmes (1982). Cigarette smoking and external locus of control among young adolescents. *Journal of Health and Social Behavior 23*(3), 253–259.

Coleman, M. and T. DeLeire (2003). An economic model of locus of control and the human capital investment decision. *The Journal of Human Resources 38*(3), 701–721.

Dohmen, T., A. Falk, D. Huffman, and U. Sunde (2006, October). The intergenerational transmission of risk and trust attitudes. IZA Discussion Paper No. 2380, Institute for the Study of Labor—IZA, Bonn, Germany.

Duncan, G. J. and J. N. Morgan (1981). Sense of efficacy and subsequent change in earnings-a replication. *The Journal of Human Resources 16*(4), 649–657.

Elliott, M. (1996). Impact of work, family, and welfare receipt on women's self-esteem in young adulthood. *Social Psychology Quarterly 59*(1), 80–95.

Evans, W. M. and R. M. Schwab (1995). Finishing high school and starting college: Do catholic schools make a difference? *The Quarterly Journal of Economics 110*, 941–974.

Goldsmith, A. H., J. R. Veum, and W. J. Darity (1997). The impact of psychological and human capital on wages. *Economic Inquiry 35*(4), 815–829.

Gottschalk, P. (2005). Downward nominal-wage flexibility: real or measurement error? *The Review of Economics and Statistics 87*(3), 556–568.

Haveman, R. and B. Wolfe (1995). The determinants of children's attainments: A review of methods and findings. *Journal of Economic Literature 33*, 1829–1878.

Heckman, J. J., J. Stixrud, and S. Urzua (2006, July). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics 24*(3), 411.

Ku, I. and R. Plotnick (2003). Do children from welfare families obtain less education? *Demography 40*(1), 151–170.

Larum, J. and J. J. Beggs (1989). What drives Australian teenage labour force participation? *The Australian Journal of Statistics 31A*, 125–142.

Le, A. T. and P. W. Miller (2005). Participation in higher education: equity and access? *The Economic Record 81*, 152–165.

Lillard, L. A. and C. W. Panis (2003). *aML User's Guide and Reference Manual, Version 2*. Los Angeles, California, USA: EconWare.

Linz, S. J. and A. Semykina (2008). Attitudes and performance: An analysis of Russian workers. *The Journal of Socio-economics 37*, 694–717.

Marks, G., N. Fleming, M. Long, and J. McMillan (2000). Patterns of participation in Year 12 and higher education in Australia: Trends and issues, Longitudinal Surveys of Australian Youth. Research Report 17, Australian Council for Educational Research, Melbourne, Australia.

Marks, G. N. (2004). School sector differences in tertiary entrance: improving the educational outcomes of government school students. *Australian Social Monitor 7*(2), 43–47.

Marks, G. N., J. McMillan, and K. Hillman (2001). Tertiary entrance performance: the role of student background and school factors. Research Report 22, Australian Council for Educational Research, Longitudinal Surveys of Australian Youth, Melbourne, Australia.

Murnane, R. J., J. B. Willett, M. J. Braatz, and Y. Duhaldeborde (2001). Do different dimensions of male high school students skills predict labor market success a decade later? Evidence from the NLSY. *Economics of Education Review 20*, 311–320.

Osborne-Groves, M. (2005). How important is your personality? Labor market returns to personality for women in the US and UK. *Journal of Economic Psychology 26*, 827841.

Pearlin, L. I. and C. Schooler (1978, Mar.). The structure of coping. *Journal of Health and Social Behavior 19*(1), 2–21.

Ray, J. (1984). Some pitfalls in measuring perceived locus of control as an economic variable: A comment on Buchele. *The Journal of Human Resources 19*(4), 580–581.

Rich, L. M. (1999). Family welfare receipt, welfare benefit levels, and the schooling and employment status of male youth. *Social Science Research 28*, 88–109.

Rotter, J. B. (1972). Generalized expectancies for internal versus external control reinforcement. In J. B. Rotter, J. E. Chance, and E. J. Phares (Eds.), *Applications of a Social Learning Theory of Personality*, pp. 260–295. New York, USA: Holt, Rinehart and Winston, Inc.

Rotter, J. B. (1990). Internal versus external control of reinforcement: A case history of a variable. *American Psychologist 45*(4), 489–493.

Tables and Figures

Table 1.—Locus of control and the educational outcomes of the Youth in Focus (YIF) cohort in Australia. Association between measures of locus of control and youths' educational outcomes.

	Year 12 Completion			Obtained University Entrance Score			ENTER Score		
Locus of Control variable[a]	Mean Graduates	Mean Dropouts	Difference in Means[b] *(p-value)*	Mean for those who obtained ENTER	Mean for those who did not obtained ENTER	Difference in means[b] *(p-value)*	Mean ENTER those who Agree (Str.)	Mean ENTER those who Disagree (Str.)	Difference in means[b] *(p-value)*
I cannot solve some of my problems	.183	.246	.000	.682	.696	.599	74.93	73.58	.256
	(.387)	(.431)		(.466)	(.460)		(18.02)	(17.18)	
I feel being pushed around in life	.392	.418	.195	.673	.707	.105	73.50	74.01	.593
	(.488)	(.493)		(.469)	(.455)		(16.73)	(17.71)	
I have no control over things happenning to me	.139	.197	.000	.642	.701	.054	71.22	74.20	.030
	(.346)	(.398)		(.480)	(.458)		(17.05)	(17.34)	
I can do anything I set my mind to	.915	.875	.001	.700	.626	.056	73.85	73.43	.810
	(.279)	(.331)		(.458)	(.485)		(17.12)	(19.94)	
I feel helpless in dealing with my problems	.286	.376	.000	.672	.702	.195	73.30	73.99	.500
	(.452)	(.485)		(.470)	(.458)		(16.91)	(17.50)	
What happens to me mostly depends on me	.940	.918	.037	.696	.650	.305	73.79	74.35	.779
	(.237)	(.274)		(.460)	(.479)		(17.37)	(16.82)	
There is little I can do to change things in my life	.136	.210	.000	.613	.706	.003	72.18	74.05	.186
	(.343)	(.408)		(.488)	(.456)		(17.53)	(17.31)	

Notes: Standard errors in parentheses.

[a] From the original locus of control variables I created indicators that take value one if individuals answered that they *strongly agree* or *agree* with the statement in the question, and value zero if they *disagree* or *strongly disagree*.

[b] The column reports p-values for a test of equality of means with unequal variances.

Source: Youth in Focus data wave 1.

Table 2.—Measures of parental socio-economic background (disadvantage) and educational outcomes of the Youth in Focus (YIF) cohort in Australia.

	Year 12 Completion		Obtained University Entrance Score		ENTER Score	
Measures of Parental Disadvantage	*Proportion who graduated*	*Equality with group A*$^{(a)}$ *(p-value)*	*Mean*	*Equality with group A*$^{(a)}$ *(p-value)*	*Mean ENTER Score*	*Equality with group A*$^{(a)}$ *(p-value)*
A: Youth unexposed to disadvantage	.776	.000$^{(b)}$	.768	.000$^{(b)}$	75.05	.000$^{(b)}$
	(.417)		(.422)		(17.14)	
B: Youth exposed to six or more years of disadvantage	.549	.000	.537	.000	69.43	.000
	(.498)		(.499)		(18.42)	
C: Youth exposed to disadvantage when aged 10+ years	.686	.000	.684	.004	71.64	.010
	(.464)		(.465)		(18.09)	
D and F: Youth exposed to disadvantage when aged 6-10 years	.684	.000	.658	.000	72.89	.075
	(.465)		(.475)		(16.80)	
E: Youth exposed to disadvantage when aged less than 6 years	.640	.000	.628	.000	72.65	.112
	(.481)		(.484)		(16.79)	

Notes: Standard errors in parentheses.
$^{(a)}$ Except for the first row, this column shows the p-value for a test of equality of means with unequal variance between the group in the corresponding row and the non-disadvantaged group in the first row.
$^{(b)}$ P-value for the comparison of means with unequal variances between the non-disadvantaged group (Stratum A) and *any* degree of disadvantage (Strata B, C, D and F, and E).
Source: Youth in Focus data wave 1.

Table 3.—Association between youths' locus of control and parental socio-economic background variables of the Youth in Focus (YIF) cohort in Australia.

Question: *How strongly do you agree or disagree with the following statements:*	All Groups Strongly Agree or Agree (%)	Strongly Agree or Agree for Highly Disadvantaged and Non-disadvantaged Youths (%)	
		Stratum B	Stratum A
1. There is really no way I can solve some of the problems I have	20.3	24.6	19.7
2. Sometimes I feel that I'm being pushed around in life	40.0	43.2	36.8
3. I have little control over the things that happen to me	15.7	17.9	13.4
4. I can do just about anything I really set my mind to	90.3	88.1	91.2
5. I often feel helpless in dealing with the problems of life	31.4	34.6	29.1
6. What happens to me in the future mostly depends on me	93.3	92.6	94.5
7. There is little I can do to change many of the important things in my life	15.9	18.2	12.7

Source: Youth in Focus data wave 1.

Table 4.—Starting school age rules for members of the Youth in Focus (YIF) cohort. Member of the YIF cohort were born between 1 October 1987 and 31 March 1988.

State	*Years of Education in State*	*School Starting Age Rule*	*Year of Starting School*	*Finish Year 12 (if normal progression)*
Australian Capital Territory	Kindergarten + 12 years	Start if aged 5 years by 30th April in the year of commencement	1993	2005
New South Wales	Kindergarten + 12 years	Start if aged 5 years by 31st July in the years of commencement. If born after March tend to hold over.	1993	2005
Northern Territory	12 years	Start if 5 years old by start of school year	1994 (born Oct. 87-Jan. 88) and 1995 (born Mar. 88)	2005 (born Oct. 87-Jan. 88) and 2006 (born Mar. 88)
Queensland	12 years	Start if 5 years old by the end of previous year	1994 (born Oct.-Dec. 87) and 1995 (born Jan.-Mar. 88)	2005 (born Oct.-Dec. 87) and 2006 (born Jan.-Mar. 88)
South Australia	Kindergarten + 12 years	Start if aged 5 years by 31st May in the year commencement	1993	2005
Tasmania	Kindergarten + 12 years	Start if aged 5 years by 31st June in the year of commencement	1993	2005
Western Australia	12 years	Start if 5 years old by the end of previous year	1994 (born Oct.-Dec. 87) and 1995 (born Jan.-Mar. 88)	2005 (born Oct.-Dec. 87) and 2006 (born Jan.-Mar. 88)
Victoria	Kindergarten + 12 years	Start if aged 5 by 30th April in the year commencement	1993	2005

Note: For the model, the Northern Territory is collapsed with Queensland and Western Australia because there are only a few observations from this state.

Table 5.—Probability of completing Year 12 for youths who were born (i) Oct-Dec 1987 and (ii) Jan-Mar 1988. Age starting school rules only apply to those YIF youth in Queensland and Western Australia.

	Probability of Year 12 completion at survey interview for those who were:		*Difference in probability*
	Born Oct.-Dec. 1987	*Born Jan.-Mar. 1988*	*(p-value)*
States in which rule applies to YIF cohort:			
Queensland	.766	.684	.082
	(.424)	(.466)	(.008)
Western Australia	.726	.650	.076
	(.447)	(.478)	(.096)
States in which rule does not applies to YIF cohort:			
Australian Capital Territory	.765	.696	.069
	(.431)	(.470)	(.577)
New South Wales	.677	.605	.072
	(.468)	(.489)	(.008)
South Australia	.626	.594	.032
	(.485)	(.493)	(.561)
Tasmania	.500	.253	.247
	(.505)	(.437)	(.005)
Victoria	.731	.570	.161
	(.444)	(.496)	(.000)

Notes: Students, youths who were born early in the cohort (Oct. to Dec. 1987) in any other state also show higher probabilities of graduation. Observations from the Northern Territory are included with Western Australia because there are only a small number of observations from this. Standard deviations in parentheses, except for the third column where p-values for the test of means are in parentheses.

Table 6.—The α-parameter estimates for the measurement part of the model. All parameters in the models, including the ones reported in this table, are estimated jointly by maximum likelihood.

Measurement Model's Variables[(a)]	*High-School Completion*	*Obtained ENTER Score*	*ENTER* Score
I cannot solve some of my problems (α_1)	1.00[(b)]	1.00[(b)]	1.00[(b)]
I feel being pushed around in life ($\hat{\alpha_2}$)	.903	.904	.897
	(.065)	(.065)	(.068)
I have no control over things happenning to me ($\hat{\alpha_3}$)	1.150	1.150	1.147
	(.091)	(.090)	(.092)
I can do anything I set my mind to ($\hat{\alpha_4}$)	$-.576$	$-.576$	$-.576$
	(.047)	(.047)	(.048)
I feel helpless in dealing with my problems ($\hat{\alpha_5}$)	1.107	1.110	1.108
	(.079)	(.079)	(.81)
What happens to me mostly depends on me ($\hat{\alpha_6}$)	$-.403$	$-.403$	$-.403$
	(.044)	(.044)	(.043)
There is little I can do to change things in my life ($\hat{\alpha_7}$)	.671	.673	.673
	(.052)	(.052)	(.051)
$\hat{\sigma}_\ell$ [(c)]	1.020	1.018	1.021
	(.053)	(.053)	(.053)

Notes: Heteroscedasticity-robust Standard errors in parentheses.
[(a)] The locus of control variables take four values: 1 if *strongly agree*, 2 if *agree*, 3 if *disagree*, and 4 if *strongly disagree*. The interpretation of the locus of control latent variable, therefore, corresponds to higher values (positive) for internals and low values (negative) for externals.
[(b)] Set to 1.
[(c)] Estimated standard deviation for the latent locus of control, LC^*. $LC^* \sim N(0, \sigma^2_\ell)$.

Table 7.—The determinants of youths' educational outcomes in Australia, Probit marginal effects.

Dependent Variable:	*Probit Model:* Year 12 Completion	*Probit with Selection:* Obtained ENTER Score	*Censored Regression:* ENTER Score
Locus of Control (higher values for *internals*)[(a)]	.061*	.076**	.95*
Exposure to disadvantage when growing up			
B: Youth exposed to six or more years of disadvantage	−.099***	−.198***	−4.29***
C: Youth exposed to disadvantage when aged 10+ years	−.029	−.037	−.96
D and F: Youth exposed to disadvantage when aged 6-10 years	−.080***	−.106***	−1.24
E: Youth exposed to disadvantage when aged less than 6 years	−.050	−.116***	−1.68
Other characteristics			
Living with both parent at 14 years old	.065***	.012	−.66
Male	−.120***	−.066**	−3.79
Indigenous	−.170**	−.225**	−7.30
Parents read at night when young	.016	.064**	3.30**
Parents helped with homework	−.024	−.063**	3.34**
Father has a degree	.088***	.048	1.11
Mother has a degree	.036*	.070**	1.10
Mother completed Year 12	.054**	.023	4.93***
Father completed Year 12	.025	.093***	3.70***
Either parent is immigrant–non-English speking background	.118***	.079**	2.77**
Either parent is immigrant –English speaking background	−.019	−.044	1.74
Early born (Oct.-Dec. 1988) in Queensland or Western Australia	.036**		
Early born (Oct.-Dec. 1988) in other states	.167***		
Observations	2065	1506	1105

Notes: ***, **, and * denote significance at 1%, 5%, and 10% of the underlying coefficient. See Appendix Table A3 for parameter estimates.

[(a)] The marginal change is from someone with a locus of control at the 25th percentile to someone who has a locus of control at the 75th of the scale. A one standard deviation change (i) increases the probability of Year 12 completion by 2.2 percentage points and (ii) the probability of obtaining an ENTER score by 2.9 percentage points.

Appendix

Table A1.—Source of measure of parental disadvantage/socio-economic background.

Stratum Code	Stratification Category	Proportion in Admin. Data
A	No parental welfare history	40.9%
B	Heavy exposure to welfare programs (more than six total years on income support)	27.5%
C	First exposure to the welfare system after 1998	8.5%
D	First exposure to the welfare system between 1994 and 1998 and less than three total years on welfare	8.5%
E	First exposure prior to 1994 and less than six total years on welfare	9.5%
F	First exposure to the welfare system between 1994 and 1998 and more than three but less than six total years on welfare	5.1%

Notes: The original stratification based on parental income support histories in the Youth in Focus (YIF) data. With some modifications these categories constitute the measure of parental disadvantage used in this paper's analysis. See Figure A1 for the modifications.
Source: Breunig et al. 2007

Table A2.—Variable definition and sample descriptive statistics.

Variable definition	*Mean*	*Std.Dev.*	*Min.*	*Max*	*Obs.*
=1 if individual had completed Year 12 at interview, =0 otherwise	.717	(.451)	0	1	3723
University entrance score, 0 for those who did not take it	50.348	(33.938)	0	99.98	2294
I cannot solve some of my problems	3.044	(.753)	1	4	2705
I feel being pushed around in life	2.751	(.811)	1	4	2702
I have no control over things happenning to me	3.131	(.709)	1	4	2699
I can do anything I set my mind to	1.756	(.657)	1	4	2696
I feel helpless in dealing with my problems	2.834	(.781)	1	4	2698
What happens to me mostly depends on me	1.636	(.644)	1	4	2702
There is little I can do to change things in my life	3.125	(.703)	1	4	2701
A: Youth unexposed to disadvantage	.252	(.434)	0	1	3723
B: Youth exposed to six or more years of disadvantage	.365	(.482)	0	1	3723
C: Youth exposed to disadvantage when aged 10+ years	.127	(.333)	0	1	3723
Stratum D	.102	(.302)	0	1	3723
Stratum F	.054	(.226)	0	1	3723
E: Youth exposed to disadvantage when aged less than 6 years	.100	(.300)	0	1	3723
Regional dummy variables					
ACT	.013	(.115)	0	1	3723
VIC	.250	(.433)	0	1	3723
WA	.103	(.305)	0	1	3723
NT	.003	(.057)	0	1	3723
QLD	.218	(.413)	0	1	3723
SA	.078	(.269)	0	1	3723
TAS	.026	(.159)	0	1	3723
Characteristics					
=1 if youth is male, 0 otherwise	.468	(.499)	0	1	3723
=1 if youth is indigenous, 0 otherwise	.040	(.197)	0	1	3714
=1 if parent read to youth at night, 0 otherwise	.471	(.499)	0	1	2715
=1 if parent helped youth with homework, 0 otherwise	.554	(.497)	0	1	2717
=1 if youth lived with both parents at age 14, 0 otherwise	.664	(.472)	0	1	3711
=1 if youth's father had a degree when the youth was aged 14, 0 otherwise	.539	(.499)	0	1	3188
=1 if youth's mother had a degree when the youth was aged 14, 0 otherwise	.471	(.499)	0	1	3133
=1 if youth's mother was a high school graduate when youth was 14, 0 otherwise	.469	(.499)	0	1	3446
=1 if youth's father was a high school graduate when youth was 14, 0 otherwise	.409	(.492)	0	1	3446
=1 if either parent is immigrant from non-English-speaking country, 0 otherwise	.246	(.430)	0	1	3723
=1 if either parent is immigrant from English-speaking country, 0 otherwise	.164	(.370)	0	1	3723

Source: Author's calculations based on data from Youth in Focus (YIF) data, wave 1.

Table A3.—Probit estimates of Year 12 Completion and probit model of Obtaining ENTER Score with selection.

	Probit Model: High-School (HS) Completion	*Probit with Selection: Obtained University Entrance Score*	
Variables		*HS*	*ENTER*
Locus of Control (higher values for *internals*)	.059	.057	.085
	(.036)	(.035)	(.040)
B: Youth exposed to six or more years of disadvantage	−.316	−.319	−.536
	(.092)	(.092)	(.099)
C: Youth exposed to disadvantage when aged 10+ years	−.100	−.100	−.111
	(.110)	(.110)	(.116)
D and F: Youth exposed to disadvantage when aged 6-10 years	−.261	−.258	−.300
	(.098)	(.098)	(.107)
E: Youth exposed to disadvantage when aged less than 6 years	−.167	−.165	−.326
	(.119)	(.118)	(.125)
Living with both parent at 14 years old	.201	.203	.033
	(.075)	(.075)	(.092)
Male	−.373	−.374	−.178
	(.063)	(.063)	(.082)
Indigenous	−.475	−.484	−.577
	(.191)	(.190)	(.253)
Parents read at night when young	.049	.052	.173
	(.067)	(.067)	(.074)
Parents helped with homework	−.077	−.074	−.173
	(.067)	(.067)	(.074)
Father has a degree	.275	.276	.130
	(.068)	(.068)	(.080)
Mother has a degree	.114	.115	.189
	(.068)	(.068)	(.074)
Mother completed Year 12	.172	.169	.062
	(.069)	(.070)	(.081)
Father completed Year 12	.078	.073	.253
	(.071)	(.071)	(.077)
Either parent is immigrant–non-English speking background	.401	.401	.218
	(.080)	(.079)	(.087)
Either parent is immigrant –English speaking background	−.058	−.069	−.117
	(.084)	(.084)	(.095)
Policy States (QLD, WA, NT)	.539	.559	
	(.096)	(.093)	
Interaction (Early Born × Policy States)	−.277	−.229	
	(.144)	(.148)	
Early born (October to December, 1987)	.421	.368	
	(.073)	(.093)	
Constant	.158	.172	.327
	(.121)	(.123)	(.213)
$\hat{\rho}$			.551
			(.240)
Observations	2065	2065	1506

Notes: Results for the Censored Regression Model of ENTER scores are presented in Table 7 and are directly interpretable. For each outcome, parameters are estimated jointly by maximum likelihood (including parameter in the measurement part of the model). See Table 6 for parameter estimates in the measurement part of the models. *Source:* Author's calculations based on data from Youth in Focus (YIF) data, wave 1.

Table A4.—Testing for differences in the effect of parental disadvantage on youths' educational outcomes by youths' age at disadvantage. Likelihood Ratio ests for hypotheses related to parental disadvantage in the baseline model discussed in the text.

	Year 12 Completion Model		*Obtained an ENTER Score*		*ENTER Scores*	
Hypothesis	*LR-statistic*	*P-value* (χ^2)	*LR-statistic*	*P-value* (χ^2)	*LR-statistic*	*P-value* (χ^2)
Does disadvantage matter for outcomes?						
$H_0 : \theta_B = \theta_C = \theta_{DF} = \theta_E = 0$	13.98	(.007)	29.72	(.000)	11.05	(.025)
Does the effect of being highly disadvantaged matter differently from the effect of being intermediately disadvantaged?						
$H_0 : \theta_B = \theta_C$	4.28	(.038)	13.58	(.000)	6.82	(.009)
$H_0 : \theta_B = \theta_{DF}$	0.32	(.567)	4.80	(.028)	3.77	(.527)
$H_0 : \theta_B = \theta_E$	1.64	(.199)	2.72	(.099)	2.38	(.123)
Does youth's age at the time of disadvantage affect educational outcomes differently?						
$H_0 : \theta_C = \theta_{DF}$	1.92	(.166)	2.18	(.140)	2.01	(.156)
$H_0 : \theta_{DF} = \theta_E$	0.24	(.625)	2.22	(.135)	1.93	(.165)
$H_0 : \theta_C = \theta_E$	0.73	(.396)	1.01	(.315)	1.13	(.288)

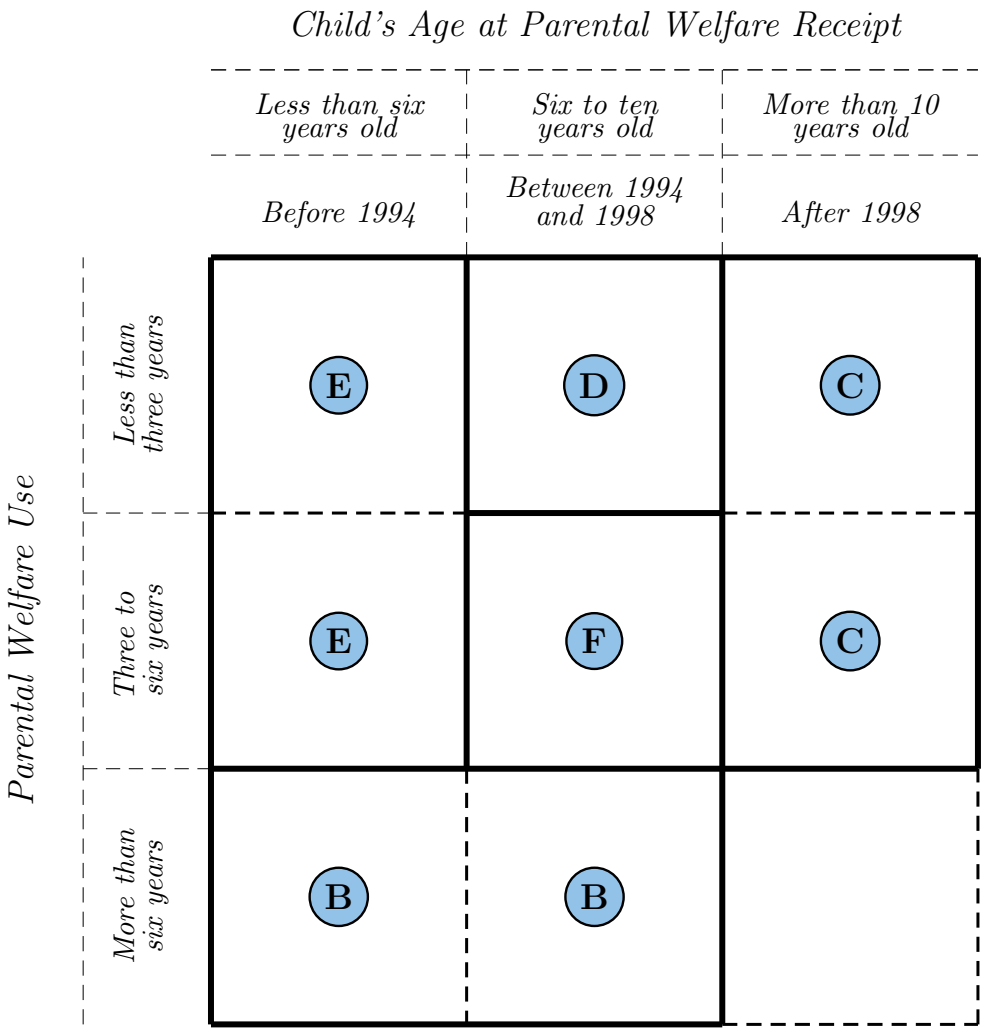
Notes: Restricted models are not shown but are available upon request. P-values in parentheses. The degrees of freedom are four for the test in the first row and one for all other tests.

Table A5.—The effect of disadvantage at home on youths' locus of control. The figures come from a model in which latent locus of control is regressed on disadvantage variables.

Dependent Variable: Latent Locus of Control	*Model With No Controls*		*Model With Controls*	
Disadvantage Variable	*Coefficient*	*S.E.*	*Coefficient*	*S.E.*
B: Youth exposed to six or more years of disadvantage	−.150	(.065)	−.116	(.073)
C: Youth exposed to disadvantage when aged 10+ years	−.081	(.084)	−.038	(.085)
D and F: Youth exposed to disadvantage when aged 6-10 years	.029	(.075)	.056	(.074)
E: Youth exposed to disadvantage when aged less than 6 years	−.010	(.101)	.031	(.103)
Observations	2065		2065	

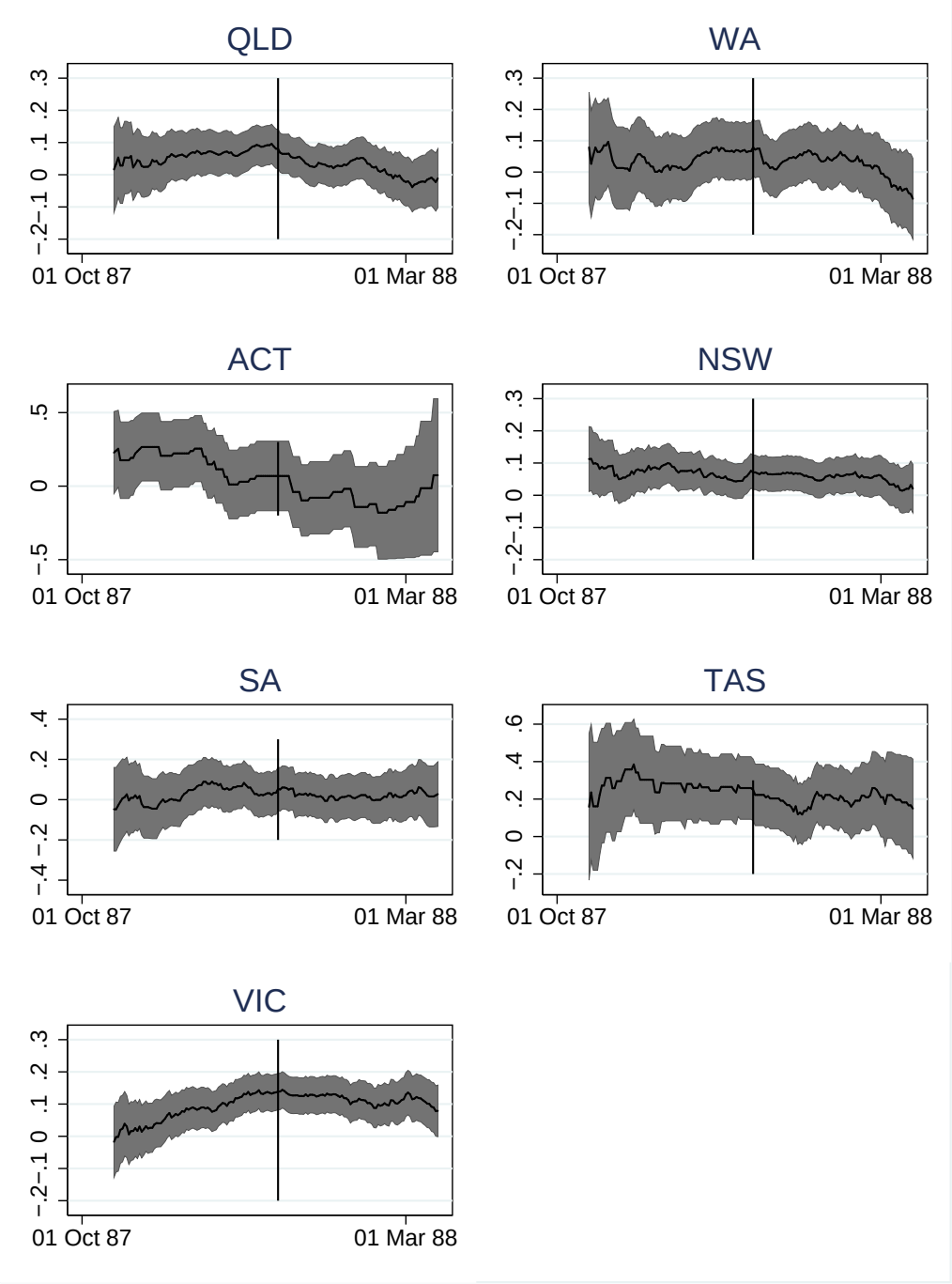
Note: Regressor include gender, parental education, parent's non-pecuniary, indigenous indicator, parental immigration status.

Figure A1.—Diagram of the stratification in the Youth in Focus (YIF) data.



Notes: The measures of disadvantage used in this paper's analysis combines groups D and F into a single category and maintains the other categories unaltered.

Figure A2.—Running difference in the probability of graduation before and after each date. The vertical lines are at 31 December 1987. These lines only apply as a rule for starting school for Queensland and Western Australia. The vertical axis represents the probability.



Notes: The data for this figure is calculated as follows: For each day between 15 October 1987 and 15 March 1988, I calculate the probability of Year 12 completion at the time of the interview for (i) those youths who were born before that day, and (ii) those youths who were born after that day. The shaded region represent normal-based confidence intervals at the 95 per cent level.