Final Report

The Impact of Middle-Years Experiences on Educational and Labour Market Outcomes

Deborah Cobb-Clark, Sonja Kassenboehmer, Trinh Le, Duncan McVicar and Rong Zhang

Melbourne Institute of Applied Economic and Social Research

Acknowledgements

This research was commissioned by the Australian Government Department of Education, Employment and Workplace Relations (DEEWR) under the Social Policy Research Services Agreement (2010--12) with the Melbourne Institute of Applied Economic and Social Research. The data used for this research come from the Youth in Focus Project which is jointly funded by DEEWR, the Australian Government Department of Families, Housing, Community Services and Indigenous Affairs and the Australian Research Council (Grant Number LP0347164) and carried out by the Australian National University. The views expressed in this report are those of the authors alone and do not necessarily represent the views of DEEWR, the Australian Government, or the Melbourne Institute. We would like to thank Anastasia Sartbayeva and Chris Ryan for sharing their programming code, Mike Helal for help with data on academic disruptions and Guyonne Kalb and DEEWR staff for comments on previous drafts. Any remaining errors are those of the authors.

July 2012

THE UNIVERSITY OF MELBOURNE

MELBOURNE INSTITUTE® of Applied Economic and Social Research

Table of Contents

Executive Summary

Background

The 'middle years' of schooling are generally accepted to cover the school years from Years 5 to 8, which are the last two years of primary schooling and the first two years of secondary schooling in most States and Territories in Australia. 'Middle years' students are thus aged around 9-14 years, broadly corresponding to what is termed the 'early adolescence' phase (ages 10-14) in developmental psychology. This period also corresponds roughly to the 'middle school' or 'middle grades school' in the United States and United Kingdom.

During the middle years students typically switch from a small, cosy primary school to a larger, more formal secondary school, and experience associated changes in teachers, friends, learning environments and learning methods. During this period young people also face the biological transformations of puberty and the psychological shifts that accompany the emergence of sexuality. For these and other reasons, the middle years can be a particularly challenging phase in young people's development.

Outline of the Research

This report examines the associations between two particular middle-years experiences – *marijuana use* and *being suspended or expelled from school* – and the educational and labour market outcomes of individuals when they reach 20 years of age. To do so the research exploits a rich source of data on Australian young people and their parents – the Youth in Focus (YIF) study – which has not previously been used for this purpose. The outcomes considered are Year 12 completion, obtaining a university entrance score, the university entrance score obtained, and whether the young person is in employment, education, or neither at age 20.

The report also goes further than the majority of existing studies in this area by explicitly examining the extent to which the associations between these middle-years experiences and later outcomes can be interpreted as *causal relationships*. In other words, the report assesses the extent to which outcomes at age 20 are different for those young people who had the middle-years experience compared to those who did not *because of the middle-years experience*, as opposed to being different for those who had the middle-years experience for other reasons.

This distinction is potentially important for policy. For example, a negative causal relationship between middle-years marijuana use and Year 12 completion implies that policy

interventions aimed at reducing marijuana use among this age group will have additional beneficial effects on Year 12 completion. A non-causal relationship implies no such positive knock-on effect, although policy makers may still want to target interventions aimed at increasing Year 12 completion rates at marijuana users if marijuana users are a group with low completion rates.

Further, the report assesses the extent to which the strength of any associations between middle-years experiences and later educational and labour market outcomes is related to the socio-economic status (SES) of the young person. Of particular interest here is whether any longer-term detrimental impacts of the two middle-years experiences considered here are stronger for low-SES young people, e.g. because of a relative lack of support in overcoming potential negative consequences of these experiences and behaviours.

Key Findings

Comparing mean outcomes between young people in the YIF study who do/ do not report the two middle-years experiences shows up some very strong associations. All of the associations with educational outcomes are negative. For example, respondents who report using marijuana by the age of 14 and respondents who report having been suspended or expelled from school have lower Year 12 completion rates than their counterparts (by 26.5 percentage points and 31.6 percentage points respectively). They also have lower probabilities of obtaining a university entrance score (by 6.4 percentage points and 16.7 percentage points respectively), lower entrance scores when a score is obtained (by 5.5 points and 7.8 points respectively), and lower probabilities of studying at age 20 (by 17.3 percentage points and 21.9 percentage points respectively). The lower probabilities of studying at age 20 correspond with higher probabilities of being in employment or neither employment nor education at age 20.

The strength of these associations tends to decrease when observable differences in the characteristics of those respondents that do/ do not have the middle-years experiences and their families are controlled for in a multivariate regression framework. In other words, part of the gaps in outcomes can be explained by observable differences in individual and family characteristics between those young people that do and do not report each middle-years experience. But there are still many strong associations, and all associations with educational outcomes remain negative. After controlling for observable differences in this way, young people who report using marijuana by the age of 14 and young people who report having

been suspended or expelled from school still have lower Year 12 completion rates than their counterparts (by 10 percentage points and 18.3 percentage points respectively), lower probabilities of obtaining a university entrance score (by 8.2 percentage points and 6 percentage points respectively), lower entrance scores when a score is obtained (by 3 points and 5.8 points respectively), and lower probabilities of studying at age 20 (by 6.3 percentage points and 19.6 percentage points respectively).

In the same way that observable differences in characteristics between those that do/ do not report the middle-years experiences can explain part of the overall gaps in outcomes, so too can *unobservable* differences in characteristics between the different groups (e.g. differences in attitudes to risk between the different groups of young people). Because such differences are unobservable we cannot control for them in a multivariate regression framework. But we can make informal (but informed) assessments of their likely importance.

Such assessments strongly suggest that none of the remaining associations between respondent suspension/expulsion and later outcomes are likely to reflect causal impacts of suspension/expulsion on the outcomes. In other words, all associations can be plausibly explained by differences in unobservable factors between those that do and those that do not report suspension/expulsion. For marijuana use the picture is more mixed. For Year 12 completion and studying at age 20 there is little evidence here to suggest causal impacts. But causal effects cannot be ruled out from middle-years marijuana use on the probability of obtaining a university entrance score and on the entrance score obtained. The evidence presented here suggests that middle-years marijuana use amongst YIF respondents reduces the probability of obtaining an entrance score by at least 8.2 percentage points, and the entrance score obtained by at least 3 points.

Evidence is also presented of negative associations between family SES and the young person's educational outcomes for the YIF sample. Using family income-support history to proxy for SES, the report shows that low-SES young people (those with a record of intensive family income-support receipt) are 23.1 percentage points less likely to complete Year 12 compared to high-SES young people (those with a record of no family income-support receipt). Similarly, the probability of obtaining an entrance score is 14 percentage points lower, the entrance score obtained is 5.3 points lower, and the probability of studying at age 20 is 19.3 percentage points lower.

There is also some evidence that associations between middle-years marijuana use and educational and labour market outcomes are stronger for low-SES individuals than for high-SES individuals, and that associations between suspension/expulsion and education and labour market outcomes are stronger for high-SES individuals than for low-SES individuals.

Policy Implications

To the extent that these findings from the YIF study generalise to the wider population of young people in Australia, the research suggests the following policy implications:

First, interventions aimed at reducing the proportion of young people who use marijuana in the middle years are unlikely to have knock-on effects on Year 12 completion rates, but may have positive knock-on effects on the proportion of Year 12 completers who obtain a university entrance score, and on the entrance score obtained.

Second, the evidence presented here suggesting no causal effect between suspension/expulsion and an individual's educational outcomes, on average, may be useful for school principals taking difficult decisions regarding the use of such sanctions as disciplinary measures of last resort.

Third, even in the absence of causal effects, the strong associations found in the data between these middle-years experiences and educational outcomes at age 20 suggest that those who tried marijuana in the middle years and those who have been suspended or expelled from school may be important groups to target for interventions intended at improving young people's educational outcomes. Targeting interventions at those suspended/expelled is likely to be more practical, however, given that data on marijuana use are unlikely to be routinely available.

1. Introduction

An extensive range of literature in disciplines such as psychology, economics, sociology, and public health outlines the critical importance of both early childhood and adolescence in shaping the life chances of individuals. Investments made during early childhood provide the bedrock for children's cognitive and non-cognitive development, educational achievement, and emotional wellbeing in later years (Heckman, 2000, 2006, for example). Young people can then go on to make critical (and increasingly independent) decisions regarding their own health, education, employment, and family arrangements during adolescence (Gruber, 2000, for example). But the period *between* early childhood and adolescence – the 'middle years' – is also a key stage in the life course. For example, educators have for decades pointed to the particular challenges associated with educating children in this age range (e.g. Juvonen et al., 2004).

These 'middle years' of schooling are generally accepted to cover the school years from 5 to 8, which are the last two years of primary schooling and the first two years of secondary schooling in most states and territories in Australia.[1] 'Middle-years' students are thus aged around 9-14, broadly corresponding to what is termed the 'early adolescence' phase (ages 10-14) in developmental psychology.[2] This period also corresponds roughly to the 'middle school' or 'middle grades school' in the United States (US) and United Kingdom (UK).

During these middle years students typically move from small, self-contained primary school classrooms to larger, more integrated learning environments in secondary schools. Secondary-school classes are typically organised around subject areas leaving students to negotiate their way through multiple classrooms with multiple teachers and a much larger set of peers. This is also a period in which young people generally face the biological transformations of puberty and the psychological shifts that accompany the emergence of sexuality. This critical life phase can be problematic for many young people as they undergo so many changes simultaneously (Eccles, 1999). However, psychologists have demonstrated that the ability to regulate intense emotions, distinguish feelings from facts, reason, make decisions, and solve problems steadily improves throughout adolescence (Byrnes, 2003; Smetana and Turiel, 2003). Nonetheless, some adolescents may make poor choices at this

[1] See http://www.aussieeducator.org.au/education/levels/middleyears.html. Secondary schooling begins at Year 8 in Queensland, South Australia and Western Australia and at Year 7 in all other states.
[2] The Journal of Early Adolescence defines 'early adolescence' as individuals aged 10-14 (see http: //jea.sagepub.com/).

time – for example to engage in risky or disruptive behaviour – which may have long-lasting consequences for their educational performance, labour market outcomes, health, and wellbeing.

Australian public policy has begun to focus more specifically on the middle years resulting in a need for a deeper understanding of the longer-term effects of middle-years experience on outcomes later in life.[3] The objective of this project is therefore to contribute to the Australian evidence base on young people by examining the impact of two specific middle-years experiences on a selection of educational and labour market outcomes for young Australians. In particular, we examine the impacts of marijuana use and suspension or expulsion from school during the middle years on: the probability of Year 12 completion; the probability of obtaining a university entrance score; the university entrance score obtained; and whether the young person is in education, employment, or neither at 20 years of age. This report then assesses the extent to which the strength of any associations between these middle-years experiences and later outcomes is related to the socio-economic status (SES) of the young person. Of particular interest here is whether low-SES young people need additional support to overcome potential negative consequences of these experiences and behaviours.

In addressing these research questions, we take advantage of unique survey data from the Youth in Focus (YIF) project which provide detailed information about educational attainment and current activity at age 18 and 20, along with retrospective data (including data provided by their mothers) about a range of events taking place during the respondents' middle years, and administrative data (from Centrelink records), for a broadly representative sample of several thousand young people born between October 1987 and March 1988 (for more details see Breunig et al., 2009).

This report identifies three main issues for consideration. First, the direction and strength of associations between the selected educational and labour market outcomes at age 20 and middle-years marijuana use and suspension/expulsion for the YIF cohort are set out. Second, the extent to which the strength and/or direction of these associations vary by SES is assessed. Third, the report assesses the extent to which these associations can be interpreted as capturing whether outcomes at age 20 are different for those young people who had the

[3] For example, the Australian Research Alliance for Children and Youth (ARACY) held a Middle Years Symposium in November 2010 to discuss policy and practice agenda for children in the middle years. It was announced in the 2011 Budget that a national survey to measure the wellbeing of children aged 8-14 (the Social Engagement and Emotional Development Survey) would be carried out in 2012-13 (Council of Australian Governments, 2011).

middle-years experience compared to those who did not *because of the middle-years experience,* i.e. *causal relationships.*

The remainder of the report is set out as follows. Section 2 briefly reviews relevant literature. Section 3 describes the YIF data used in the study. Section 4 briefly describes the study methodology, with technical details in the Appendix. Section 5 presents and discusses the research findings with respect to associations between the selected middle-years experiences and the educational and labour market outcomes. Section 6 discusses the extent to which the associations reported in Section 5 can be interpreted as causal relationships. Section 7 shows how these associations differ by SES. Section 8 summarises the findings and briefly draws out some potential implications for policy.

2. Literature review

2.1 Links between early-adolescent environments, experiences, and later outcomes

The environments in which early adolescents live – including the family, the neighbourhood, and the school – have been shown to have considerable impact on how they adapt to the internal changes associated with early adolescence, on their sense of efficacy, self-esteem, self-reliance, and on their relationships with others.

With respect to the family, Wigfield et al. (2006) note that earlier research has typically focused on the link between family demographic characteristics (family structure, family size, parents' financial resources, parents' education, parents' occupation, community characteristics, and dramatic changes in the family's economic resources) and child outcomes. However, recent research has shown that parent's beliefs and behaviours are critical in setting a climate for the development of children's motivation by providing various activities or resources in the home environment that may provide stimulation to pursue various activities across time. Wigfield et al. (2006) also show that the strength of associations between family demographic characteristics and child outcomes can in turn depend on parents' beliefs and practices.

Parent-child relationships generally begin to change during early adolescence (Eccles, 1999). As adolescents become physically mature, they often seek more autonomy and may question family rules and roles, at times leading to conflicts with parents. They may also spend less time with parents outside the home than before, which may put them at risk as they begin to experiment with new activities.

Neighbourhoods are also an important context for adolescent development, because it is during early adolescence that young people increase exploration of neighbourhood settings and social interactions with neighbours (Lenzi et al., 2011). Lenzi et al. (2011) observe that since adolescents' exposure to the local community can often be unsupervised, they can come in contact with several risks, yet at the same time can find different opportunities for positive development, creating supportive networks with people and local organisations.

Early adolescents also navigate the transition from primary school to secondary school. Not only do early adolescents have new classmates and teachers to interact with, the learning environment at the new school is also different. While the primary school is largely child-centred, the secondary school is larger, more diverse and more subject-centred (Blyth et al., 1983). Besides, performance standards are higher, making it harder to obtain good grades (Eccles et al., 1993) and to participate in extracurricular activities (Brown, 1998).

Early adolescents also become increasingly involved with their peers. In their review of the literature, Cook et al. (2007) offer three explanations for this phenomenon. First, early adolescents turn to their friends for help in understanding and adapting to the biological changes of puberty they are experiencing. Second, early adolescents have strong social identity needs, so they spend more time with their friends; feel happier when with their friends than with family members; and worry about how their friends accept them and how they are evaluated by the local social hierarchies. Third, early adolescents start to comprehend how important their peers are to them personally as they already have the cognitive skills necessary for being more empathic and understanding.

A considerable body of evidence has shown that the environments and experiences that young people were exposed to during early adolescence are associated with a variety of outcomes later in life. For example, Sandefur et al. (1992) find that individuals who did not reside with both biological parents at age 14 are less likely to graduate from high school. Tomlinson and Walker (2010) show that parental guidance during early adolescence is associated with educational attainment and labour market status in adulthood. Kaplan et al. (2005) show that early-adolescent school-related stress is negatively associated with academic performance in later adolescence. Kasen et al. (1998) show that academic achievement, academic aspirations and learning-focussed school settings during adolescence are associated with chances of dropout, pregnancy, committing a crime, criminal conviction, antisocial personality disorder, and alcohol use during young adulthood. Renda et al. (2011) show that bullies in early adolescence are more likely to display anti-social behaviour and to

be involved with police or the courts during young adulthood. Chatterji (2006a) shows that adolescent alcohol use is associated with reductions in educational attainment measured at around age 26.

2.2 The links between early-adolescent marijuana use and later educational and labour market outcomes

Recent research suggests that marijuana[4] use in early adolescence might impact negatively on subsequent educational and labour market outcomes through a number of channels including effects on cognitive efficacy and psychological health (e.g. Scheier and Botvin, 1995; Patton et al., 2002), delinquent peer affiliations (e.g. Fergusson and Woodward, 2000), and substitution of time spent under the influence of drugs for time spent studying (van Ours and Williams, 2009).

But early-adolescent marijuana use and later educational outcomes might also be correlated for other *non-causal* reasons. The explanation for such a non-causal association is that some third factor, e.g. the lack of a supervisor at home, might drive both educational outcomes and marijuana use. Another potential explanation for non-causality is that individuals with less academic inclination may be more likely to try marijuana, and these individuals tend to have poorer educational outcomes whether they are involved in drug use or not. So a statistical association cannot always be interpreted as indicating a causal relationship.

Several existing studies present quantitative evidence of a negative association between early marijuana use and subsequent years of schooling, and although these studies tend to stop short of identifying particular causal mechanisms for such an association, some do at least present credible evidence that at least part of any association is likely to reflect a negative casual relationship between marijuana use and schooling outcomes. Most of these studies exploit either longitudinal or retrospective survey data from outside Australia, but Van Ours and Williams (2009) is an important exception using Australian data.

A number of studies use US data to estimate the sign and strength of associations between adolescent marijuana use and subsequent educational and labour market outcomes, without making any particular claim regarding causality. For example, Brook et al. (1999) use longitudinal US data on around 1,200 African American and Puerto Rican young people to estimate the association between early-adolescent marijuana use and high school graduation. They find marijuana users to be twice as likely as non-users to drop out of high school. Bray

[4] The term 'marijuana' is used throughout this study, even though some studies reviewed here use 'cannabis' to refer to the same drug.

et al. (2000) also find a positive association between marijuana use and dropping out of school using an alternative source of longitudinal survey data for 1,400 school students in the US.

Burgess and Propper (1998) examine the association between (late) adolescent marijuana use and earnings measured at age 28, using US longitudinal data for over 12,000 young people tracked over time, drawn from the National Longitudinal Survey of Youth (NLSY). They find no association between 'light' marijuana use and earnings, but strong, negative associations between 'heavy' marijuana use and earnings, particularly for black men.[5]

A 'second generation' of studies, again largely using US longitudinal data, has attempted more explicitly to identify casual relationships between adolescent marijuana use and later educational outcomes. Register et al. (2001), for example, use data for around 3,500 young people from the National Education Longitudinal Study (NELS) to show that adolescent drug use reduces subsequent years of schooling by around one year, on average. Their identification strategy exploits information on the religious affiliation of the young people to instrument for marijuana use.[6] Chatterji (2006b) also uses an instrumental variables approach with NELS data – exploiting information on school and state-level drugs policies – to try to identify the causal relationship between adolescent marijuana use and years of schooling. Their preferred model suggests that adolescent marijuana use leads to a reduction of around 0.2 years in schooling.[7]

Finally, Van Ours and Williams (2009) exploit cross-section survey data, including retrospective information on marijuana use, from the Australian National Drug Strategy Household Survey, covering almost 5,000 males and 7,000 females. They too use an instrumental variables approach – using adolescent tobacco smoking to predict marijuana use – to try to identify the casual relationship between marijuana use prior to age 15 and years of education.[8] They find that marijuana use by the age of 15 reduces years of schooling by 1.2 years on average. They also show that this effect is stronger for girls than for boys, and that

[5] Heavy use is defined as using more than 50 times in the past year. They also examine associations between adolescent marijuana use and other outcomes, including labour supply, with mixed results.

[6] The assumption behind this approach is that religion impacts on substance use but not on years of schooling, other than through any impact on substance use. Conditional on this assumption, the religion variables can be used to predict marijuana use, with the predicted values then replacing the actual values in the regression equation for years of schooling.

[7] The instrumental variables strategy is problematic in this case because the drugs policy variables are only weakly correlated with individual marijuana use. The preferred specification is therefore a standard OLS regression with extensive controls.

[8] They also use a bivariate duration analysis approach to test the robustness of their conclusions.

the (absolute) magnitude of the effect is higher the earlier that initiation into marijuana use occurs. They also find evidence, somewhat counter-intuitively at first glance, that it is *brighter* young women who try marijuana, although this is not the case for young men (where there is no correlation with academic ability).

2.3 The links between school suspension/expulsion and later educational and labour market outcomes

Suspension from school might impact negatively on later outcomes through a number of channels including impacts on self-respect, stigma among peers, increased contact with delinquent subculture, isolation from the school setting, and the loss of instructional time incurred during the suspension (see Costenbader and Markson, 1998; Morrison et al., 2001). But studies that allow us to quantify the strength of any associations between suspension/expulsion and later educational and labour market outcomes are limited. Mendez and Saunders (1981) and Mendez (2003) are exceptions, both using US data. No studies were found based on Australian data.

These studies both suggest a negative association between suspension/expulsion and later educational outcomes (specifically high school graduation rates). The strongest association is found by Mendez and Sanders (1981), who exploit US survey data on 1,300 young people to show that high-school graduation rates were around 40 percentage points lower for students that been suspended relative to those that had not been suspended. A weaker association is found by Mendez (2003), who uses US data tracking around 8,000 school pupils over a 13-year period to show that suspension is (moderately) negatively correlated with on-time graduation.[9]

If these conclusions were to carry over to the Australian context we might therefore expect a negative association between suspension/expulsion and Year 12 completion. But without examining Australian data we have no way of quantifying the strength of such an association. Nor does this small existing literature help us to quantify any associations between suspension/expulsion and the other selected educational and labour market outcomes for our cohort of interest, to assess whether any such associations can be interpreted as causal relationships, and to assess whether they vary by SES.

[9] In a related study that makes more of an attempt to identify casual relationships, Karakus et al. (2012) use NELS data for the US to show that middle-school behaviour problems (although not explicitly suspension/expulsion) impact negatively on high-school graduation but, conditional on graduation, have no significant impact on adult employment.

2.4 Differential impacts of middle-years experiences by socio-economic status

There are numerous existing studies that present evidence of associations between childhood SES and later outcomes[10], and existing studies that present evidence of associations between childhood SES and adolescent participation in risky behaviours[11]. However, there are no studies that examine whether the associations between middle-years experiences and later educational outcomes vary by SES, at least for marijuana use and suspension/expulsion.

Our conjecture is that SES may interact with our two middle-years experiences – marijuana use and suspension/expulsion – in shaping educational and labour market outcomes. For example, high-SES families may be more likely to have financial and other resources that can help to offset any negative effects of middle-years marijuana use on cognitive efficacy, psychological health, delinquent peer affiliations, or study-time. This report explores whether, in such families, middle-years marijuana use may therefore have a less detrimental impact on later educational and labour market outcomes. Similarly, are high-SES families better able to offset any effects of suspension/expulsion on an adolescent's self-respect, contact with delinquent subculture, and the loss of instructional time incurred during the suspension?

3. Data and descriptive statistics

3.1 The Youth in Focus study

This study draws on data from the YIF survey, which asks questions about family background, living arrangements, education, work, relationships, income, health, spare time, and aspirations and attitudes of young people in Australia. Individuals born between October 1987 and March 1988 who appeared in the Centrelink administrative database were randomly selected and invited to participate in the survey.

[10] For example, Burgess et al. (2001) find that family background, schooling and local area during adolescence are significantly related to earnings capacity and poverty risk during adulthood, with family factors having the strongest explanatory power. Israel and Seeborg (1998) find adolescent's maternal educational attainment to have significant indirect influences on family income during early adulthood, through intervening variables, especially the respondent's own educational attainment, welfare dependency and work experiences.

[11] For example, Cobb-Clark et al. (2012) use YIF data for Australia to show that participation in a variety of risky behaviours at age 18 (including marijuana use) is strongly associated with family SES (as proxied by the degree of family welfare dependence during childhood). Using a similar methodology to that used here in a later section (see Section 6), however, they conclude that these associations are unlikely to reflect causal relationships.

A young person has a Centrelink record in their own right if he/she received any government payment, such as Youth Allowance. They may also have a Centrelink record because while growing up their family received a payment such as Family Tax Benefit Part A or B, or any social security support, such as the Disability Support Pension. Over 98 percent of young people born between October 1987 and March 1988 in the overall Australian population appear in the Centrelink sampling frame (Breunig et al., 2009).

Respondents were first interviewed in late 2006 (wave 1, when they were around 18 years of age) and then in late 2008 (wave 2, when they were around 20 years of age). There were 4,079 respondents in wave 1 and 3,623 respondents in wave 2. Wave 2 respondents included both continuing respondents (those who participated in wave 1) and new entrants.

A particular strength of the YIF data is that one of the parents of the selected individuals, usually the mother, was also invited to answer the parent questionnaire. Parents, however, were only interviewed in wave 1.[12] It should be noted that while 4,079 youths and 3,964 parents participated in wave 1, only 2,430 youth-parent pairs could be formed. This is because many young respondents participated while their parents did not and vice versa. Wave 2 included 3,623 young respondents, 1,879 of whom had matched parental records.

In practice, the responding parent was the biological mother for 96.5 percent of young respondents.[13] In order to get consistent information on parents we therefore follow Barón and Cobb-Clark (2010) and Cobb-Clark et al. (2012) in restricting the sample to those young people with a parent questionnaire filled out by the biological mother. This excludes 50 cases. Also excluded are cases where the parent had not lived with the young respondent in the past five years (8 cases) or where the young respondent was still attending school at wave 2 (5 cases).

3.2 Key variables

3.2.1 Outcomes

We consider three educational outcomes (Year 12 completion, obtaining a university entrance score, and the entrance score itself), and one labour market outcome (whether studying, and working or neither at age 20, which is hereafter sometimes referred to as 'economic activity' for convenience). All outcomes are measured at wave 2. These important educational and

[12] Some of the new entrants to the survey in wave 2 had parents who were interviewed in wave 1. This is because the new entrants were invited to participate in wave 1 but did not participate, while their parents did.
[13] 2.1 percent of parent questionnaires were completed by the biological father with the remainder completed by a stepparent or guardian.

labour market outcomes for young people have also been analysed by earlier studies using the same data set (Barón and Cobb-Clark, 2010; Cobb-Clark et al., 2012, for example). Outcomes are measured at wave 2 because at that time (around the age of 20) even individuals who were born later in the year, entered school late or had repeated a year had enough time to have completed Year 12 and to have decided whether or not to pursue further education.

School leavers in Australia are required to have a university entrance score in order to apply for university entrance. In some states and territories, the score is derived from a state-wide exam, in others, the final results of specific subjects is utilised. In both cases, students need to apply for a score and the score is only calculated for those who have completed Year 12.

University entrance scores are known by different names in different States and Territories in Australia. The Equivalent Tertiary Entrance Rank (ENTER, used in Victoria), Universities Admission Index (UAI, used in New South Wales and the Australian Capital Territory), Tertiary Entrance Rank (TER, used in South Australia, the Northern Territory, Tasmania and Western Australia) are equivalent scores. These scores range from 1 to 100, indicating the percentile ranking of the student.[14] In 2009-2010, ENTER, UAI and TER were renamed as the Australian Tertiary Admission Rank (ATAR). However, Queensland uses a different system called the Overall Position (OP). The OP score ranges between 1 and 25, where 1 is the highest and 25 is the lowest possible score. We transform the OP score to match the other scores using the conversion factors that university administrators use when comparing Queensland school leavers with those from other states for the purpose of university admission.[15] For simplicity, hereafter the term ATAR score is used to refer to the consistent university entrance score.

The economic activity variable distinguishes three mutually exclusive categories. The first category (denoted 'studying') includes individuals who were studying only or studying and working.[16] The second category ('working') includes those who were working but not engaged in any study. The last category includes those individuals who were neither studying nor working. This consists of the unemployed and those not in the labour force but not studying.

[14] Scores under 30 are reported as being 30 to the student, so reported scores of under 30 must be data errors and we recode these to missing here.
[15] See http://www.qtac.edu.au/Applying-CurrentYr12/InterstateAdmissions.html.
[16] We initially split this group into 'studying only' and 'studying and working'. But these two groups are very similar to each other (see Appendix Table 1), and the 'studying only' group is relatively small. Therefore, we combined the two groups.

3.2.2 Middle-years experiences

This report examines two middle-years experiences in particular: marijuana use by age 14 and suspension/expulsion from school. The focus on these particular experiences reflects a combination of substantive (literature and policy-related) and pragmatic (data-related) considerations.

First, focusing on these two behaviours enables us to draw on the body of existing studies that have already established the existence of negative associations between early-adolescent marijuana use and specific later educational or labour market outcomes (see Section 2.2) and between suspension/expulsion and specific later educational outcomes (see Section 2.3), at least for the US. In other words, there is enough earlier evidence to argue that strong negative associations exist between these two experiences and the four selected educational and labour market outcomes. But significant gaps remain in the evidence base specific to Australia, to our particular selected outcomes, on the extent to which such associations reflect causal relationships, and the extent to which such associations might vary by SES.

Second, these particular behaviours are plausibly susceptible to policy intervention (and, of course, are already subject to policy intervention). For example, the prevalence of adolescent marijuana use might be influenced by measures to raise its market price or to raise awareness of its potentially harmful effects. Policy makers can also plausibly influence the use of suspensions and/or expulsions by school principals or target interventions on those suspended or expelled.

Third, while the YIF has data on many childhood experiences, in most cases it is not possible to infer whether experiences occurred during the middle years, prior to the middle years, or after the middle years. Marijuana use is the most notable exception, because individuals that reported having used marijuana were asked about the age at which they first used marijuana.[17] The YIF data do not tell us precisely when any suspension/expulsion occurred, but administrative data indicate that it is relatively common for pupils in Years 5-8 (although it is most common in Years 7-10).[18] If we are prepared to assume that any causal impact of

[17] Only 2% of individuals who report trying marijuana by the age of 14 report doing so before the age of 9, and this most likely captures measurement error. We therefore treat trying marijuana by age 14 explicitly as a middle-years behaviour.

[18] For example, administrative data from New South Wales show that students are much more likely to be suspended in Years 7-10 (77% of all long suspensions take place during these years) compared to other years. Besides, expulsions are less common than suspensions (see https://www.det.nsw.edu.au/media/downloads/about-us/statistics-and-research/ key-statistics-and-reports/long-suspension-expulsions-2010.pdf). Data from Victoria on total number of days suspended suggest that 73% of suspension days are for students in Years 7-10 and 45% for students in Years 5-8.

suspension/expulsion on later educational outcomes diminishes over time since the suspension/expulsion, then a finding of no causal relationship between suspension/expulsion and outcomes at age 20 also implies no causal relationship between suspension/expulsion specific to the middle years and later outcomes.[19]

YIF also contains information on number of schools attended and on whether the respondent had repeated a year, although again these experiences cannot be explicitly assigned to the middle years. Our assumption, however, is that changing schools is not closely correlated with the age of the child, other than for the transition between primary and secondary school. On grade repetition, data from other sources indicate that it takes place mostly in the first two years of primary school.[20] Both number of schools and grade repetition are therefore included as 'school experience' controls, but not treated specifically as middle-years experiences. We also control for family characteristics when the respondent was 14 years of age, including whether the respondent lived with both parents, and whether the respondent's mother was employed.

3.2.3 Socio-economic status

SES is generally defined as a composite indicator of family's income, education and occupation. As in Cobb-Clark et al. (2012), SES in this study is defined based on family income-support history, which is classified as follows:[21]

- families with no history of Income Support,
- families that received less than six years of support after 1998 when the respondent was older than 10 (late moderate income-support receipt),
- families that received less than six years of support, some of which occurred before 1998 when the respondent was younger than 10 (early moderate income-support receipt), and
- families that received Income Support for more than six years while the respondent was growing up (intense income-support receipt).[22]

[19] Suspension is combined with expulsion as the latter is very uncommon, and if it does occur it often accompanies the former. Only 92 respondents (2.6% of the sample) had been expelled from school, and only 10 of those had not also been suspended from school.

[20] For example, data from the Longitudinal Surveys of Australian Youth cohort 1998 suggest that of the 6.3% of respondents who reported repeating a year prior to Year 10, 54% repeated a year in either Year 1 or Year 2, and only 23% repeated a year between Year 5 and Year 8.

[21] When examining whether associations between middle-years experiences and later outcomes vary by SES, we also consider an alternative definition of SES based on parental occupational status (see Section 7).

[22] These are the existing classifications in the YIF data, as used by Cobb-Clark et al. (2012) and others.

3.2.4 Other variables

In addition to information on the particular outcomes, middle-years and other schooling experiences, information on household structure and mother's employment when the young respondent was 14 years of age, and SES, the YIF data include detailed information on the characteristics of each responding young person and their families. We broadly follow Cobb-Clark et al. (2012) in selecting a subset of variables from the available data that aims to capture the kinds of demographic and human capital factors typically found to influence educational and labour market outcomes. In doing so we are also able to exploit the particular advantage of the matched parent-child nature of the YIF study. From the young persons' questionnaire these variables include gender, an Indigenous identifier, an indicator for being born overseas and information on the source country, and area of residence. From the mothers' questionnaire additional information includes mother's age, number of children (of the mother), an indicator for parents being born overseas, whether the mother smoked, and mother's and father's highest level of educational attainment.

By exploiting this additional information in a multivariate regression framework we can essentially compare outcomes for marijuana users with non-users who are otherwise observationally equivalent, and for those suspended/expelled with those not suspended/expelled who are otherwise observationally equivalent (see Section 4). This helps us to get closer to identifying possible causal relationships behind statistical associations.

3.3 A note on sample sizes

The number of young people in the sample that can be used for analysis – sample size – varies with the particular question under examination and the method being used, and in each case the largest possible sample is used. For descriptive statistics on young people (not linked to mother's information) the full wave 2 sample of 3,623 individuals is potentially available. In practice, non-response to particular items in the questionnaire means sample size is sometimes below 3,623.[23] The same is true for regression analyses where only the young person's information is used. Where we used matched young person and mother records, however, the maximum sample size is 1,879 young people.

Sample size is further reduced where we analyse outcomes whose occurrence depends on the occurrence of another outcome. For example, since an ATAR score can only be obtained by those who have completed Year 12, data on obtaining an ATAR score is only available for

[23] For example, around a third of the respondents did not answer the question on marijuana usage.

those who completed Year 12 and answered the question on obtaining an ATAR score. Therefore, the sample size for some regressions (e.g. the effects of marijuana and parental characteristics on the ATAR score) is as low as 800 (see Appendix Table 2). One consequence of small sample sizes is that point estimates are subject to larger margins of error and hence are less reliable. As a result, point estimates are less likely to be statistically significant than if we had access to survey data with a larger sample.

3.4 Descriptive statistics and preliminary analysis of the associations between middle-years experiences and later outcomes

3.4.1 Descriptive statistics

Table 1 presents sample means for the key variables of interest: the outcomes, the middle-years experiences, other school experiences, family controls measured at the time the young person was aged 14, the set of Income Support categories to proxy for family SES, and the other young person and parental characteristics. Three columns are presented: one for the maximum possible sample, using weights constructed by the YIF team, one for the maximum possible sample in each case but unweighted, and one for the maximum possible sample where young people can be matched with mothers. Differences between the matched and unmatched samples are small in magnitude, but for some variables they are statistically significant.

Overall, 79 percent of the respondents had completed Year 12 by around age 20.[24] Among those who completed Year 12, 81 percent obtained an ATAR score, with a mean score of 75.[25] While a large proportion of respondents are studying (58 percent), a third are working and just below 9 percent are neither working nor studying.

Suspension/expulsion is relatively common among the sampled population, with almost 20 percent of all respondents reporting this experience compared to only 7.8 percent who reported trying marijuana by the age of 14.[26]

Over 40 percent of respondents come from families with no income-support history. Just over a quarter have intensive receipt (more than six years while the respondent was growing up) and just under a quarter have moderate early receipt (less than six years of support, some of which occurred before 1998 when the respondent was younger than 10).

[24] Homel et al. (2012) show that Year 12 completion rates in the YIF are very close to those for the corresponding cohort in the 2006 Census of Population and Housing.
[25] The mean ATAR score is inflated, as ATAR scores below 30 are reported as 30.
[26] Recall that a third of the respondents declined to answer the question on marijuana usage.

3.4.2 Preliminary analysis of the associations between middle-years experiences and later outcomes

Table 2 presents sample means or sample proportions for the four educational and labour market outcomes separately for respondents who did/did not use marijuana by age 14 and separately by suspension/expulsion status. In other words Table 2 shows the unconditional relationships between respondents' educational and labour market outcomes and their middle-years experiences. Outcomes are also summarised separately by grade repetition, number of schools attended, family characteristics at age 14, and family SES (as proxied by family income-support history).

Of those individuals who tried marijuana by age 14, only 61 percent completed Year 12, compared to a completion rate of 87 percent among those who did not have the experience. In other words there is a 26 percentage point difference in Year 12 completion rates between those that did and did not use marijuana by age 14. Similarly, only 54 percent of those who had been suspended and/or expelled from school completed Year 12, whereas 85 percent of those who did not have the experience completed Year 12, i.e. a 31 percentage point difference.

Compared to those without the experience, those who tried marijuana by age 14 or had been suspended and/or expelled from school are also less likely to obtain an ATAR score (gaps of 6.4 percentage points and 16.7 percentage points respectively) and have a lower ATAR score on average (gaps of 6 points and 7.8 percentage points respectively). Middle-years marijuana users and those suspended/expelled from school are also less likely to be studying at age 20 (differences of 17.3 percentage points and 21.9 percentage points respectively), and correspondingly more likely to be working or neither studying nor working.

Respondents' average outcomes are also inversely related to their family's dependence on Income Support. Young people whose family has never received Income Support are much more likely to complete Year 12 (89 percent) than those whose family has received any support (66 to 81 percent). Less than half (48 percent) of those from families with intensive income-support receipt are studying, compared to 67 percent among those from welfare-free families.

Strong relationships also exist between respondents' outcomes and their other school experiences and family characteristics when they were aged 14. For example, those who repeated a year or attended more than two schools are less likely to have completed Year 12, to have obtained an ATAR score, have lower ATAR scores (if they have a score), and are

less likely to be studying at age 20 than those who did not share the experience. In contrast, Year 12 completion, obtaining an ATAR score, and studying at age 20 are all more likely for those who lived with both parents at age 14 or whose mother was employed when the respondent was at that age relative to others. The ATAR score obtained is higher for those living with both parents at age 14 but there is no difference by mother's employment status at age 14.

Finally, marijuana use and suspension/expulsion amongst respondents are associated not only with later outcomes but also with other observable factors (see Appendix Table A3). For example, those respondents who have been suspended/expelled from school are significantly more likely to be male or from an Indigenous background, and less likely to have been born overseas or to live in a metropolitan area.[27] Early marijuana users are also more likely to have been suspended/expelled from school, to have repeated a year or to have changed schools.

4. Methods

We have already seen – from Table 2 and its accompanying discussion in Section 3.4.2 – that amongst YIF respondents middle-years marijuana use and suspension/expulsion from school are strongly associated with the probability of Year 12 completion, the probability of obtaining an ATAR score, the ATAR score obtained, and with the probability of being in study at age 20.

These associations could reflect causal influences of the middle-years experiences, i.e. outcomes at age 20 are different for those young people who had the middle-years experience *because of the middle-years experience*. But it could also be that outcomes differ for those with and without the middle-years experience because young people in the two groups differ in other ways that are themselves correlated with educational outcomes.

This distinction is potentially important for policy. For example, a negative causal relationship between middle-years marijuana use and Year 12 completion implies that policy interventions aimed at reducing marijuana use among this age group will have additional beneficial effects on Year 12 completion. A non-causal relationship implies no such positive knock-on effect, although policy makers may still want to target interventions aimed at increasing Year 12 completion rates at marijuana users as this report indicates that marijuana users are a group with low Year 12 completion rates.

[27] A metropolitan area is a 'major city of Australia' as defined by the Accessibility/Remoteness Index of Australia (ARIA).

By exploiting additional information about the young people in a multivariate regression framework, however, we can essentially compare outcomes for marijuana users with non-users who have the same observed characteristics. Similarly, we can compare outcomes for those suspended/expelled with those not suspended/expelled who have the same observed characteristics. This assists in identification of possible causal relationships behind the statistical associations presented in Table 2.

The following section (Section 5) presents and discusses the results, with a separate sub-section for each outcome. Technical details can be found in the Appendix.

The multivariate regression approach only correctly identifies a causal relationship between the middle-years experience and the outcome, however, if there are no *unobserved* differences between those with and without the experience that are also correlated with the outcome in question.[28] Finding *observed* differences between the two groups (as demonstrated by Table A3 in the Appendix) may indicate that unobserved differences exist between them. Such unobserved factors could be individual idiosyncratic characteristics (such as motivation) or family idiosyncratic characteristics (such as parental support). This is an example of a common problem in trying to identify causal relationships using non-experimental data known as *selection on unobservables*.

Section 6 of this report examines the extent to which the associations reported in Section 5 can be interpreted as causal relationships, despite the likelihood of selection on unobservables, by using a method developed by Altonji et al. (2005). This method provides an informal way of assessing the likely importance of selection on unobservables, and therefore of assessing the degree to which a statistical association might be interpretable, at least in part, as a causal effect. As before, technical details can be found in the Appendix.

Finally, to examine whether our two middle-years experiences – marijuana usage and suspension/expulsion from school – are more or less strongly associated with educational and labour market outcomes for those from low-SES families compared to those from high-SES families, we re-estimate the multivariate regression models as presented in Section 5, but splitting the sample into low-SES and high-SES parts. Although this approach allows an informal comparison of the magnitude of associations for young people from low-SES and high-SES families, it does not allow us to formally test whether any such differences are

[28] Earlier we gave the example of unobserved academic inclination, where individuals with less academic inclination may be more likely to try marijuana and less likely to achieve a particular educational outcome whether they are involved in drug use or not.

statistically significant. We therefore also re-estimate the regression models on the whole sample but including interaction terms between the SES and middle-years experiences dummies.

5. Multivariate Regression Results

This section reports the estimation results on the relationships between the various observed factors included in the regression analyses and the educational and labour market outcomes measured at age 20 for the YIF respondents. For binary outcomes (Year 12 completion and obtaining an ATAR score), probit models have been estimated, and results are presented in the form of average marginal effects which show the average change in the probability of achieving the outcome when increasing the focal explanatory variable by one unit while holding other explanatory variables constant.[29] Similar marginal effects are presented for activity at age 20, estimated using a multinomial logit model. Where the outcome is measured on a continuous scale – the ATAR score – OLS estimates are presented.

Results for each outcome are presented in a separate sub-section. A brief summary which draws the key results together is then presented.

5.1 Year 12 completion

The average marginal effects of the middle-years experiences and other observable factors on the probability of completing Year 12 are presented in Table 3. The first column presents the results from a basic model where only the individual's experiences and characteristics are included as explanatory variables. This has the advantage of a larger sample size but the disadvantage of not controlling for observable household and parental factors. The second specification adds controls for family characteristics when the respondent was 14 years of age and family income-support history, which marginally reduces sample size. In the third specification, all additional observed parental characteristics (mostly taken from the parent questionnaire) are added, which cuts the sample size by around half. The fourth column – with the same explanatory variables as the second specification but executed only on the sample with matched parental records – is included as a further robustness check.[30]

[29] The marginal effects of dummy variables are calculated as the change in predicted probability when moving from a value 0 to 1.

[30] The results from this specification (column 4) are very similar to those from the second specification (column 2), which include the same explanatory variables but are estimated on the 'full' sample (i.e. all respondents with non-missing data on the regression variables). This is consistent with evidence from Table 1 that the sample with matched parental records is, on average, similar to the full sample. Differences in estimates between the

Despite the smaller sample size, the third version of the model (column 3) is the preferred version because it offers the greatest degree of control for observable differences between those that do and do not have the middle-years experiences.[31] The following discussion therefore draws on these estimates in particular. The key conclusions regarding the association between middle-years experiences and Year 12 completion are highly robust to choice of model in any case.[32]

First consider the estimated association between middle-years marijuana use and the probability of Year 12 completion. Once other observable factors are controlled for, young people who report using marijuana in the middle years are 10 percentage points less likely to complete Year 12 than those who do not report middle-years marijuana use (see column 3). Compare this to the 26.5 percentage point difference shown in Table 2. The implication is that 16.5 percentage points of this larger difference can be explained by other observable differences between young people that do and do not use marijuana in the middle years.

Second, consider the estimated association between suspension/expulsion from school and the probability of Year 12 completion. Once other observable factors are controlled for, young people who report having been suspended/expelled from school are 18.3 percentage points less likely to complete Year 12 than those who do not report suspension/expulsion. Compare this to the 31.6 percentage point difference shown in Table 2. The implication is that 13.3 percentage points of this larger difference can be explained by other observable differences between young people that do and do not report suspension or expulsion from school.

Of the other variables, repeating school years, number of schools attended, coming from a low-SES family (proxied by intensive Income Support receipt over the young person's childhood), and number of children in the family are all (independently) negatively associated with Year 12 completion. Living in a metropolitan area, mother's and father's levels of educational attainment are all positively associated with Year 12 completion.

---

second and third model specifications (columns 2 and 3) can therefore be attributed to model differences rather than sample differences.

[31] This is reflected in a higher r-squared, which indicates that the model explains a higher proportion of the variation in Year 12 completion across individuals than the alternative versions of the model.

[32] This suggests that the middle-years experiences are little correlated with family characteristics, family income-support history and parental characteristics. Note, however, that by adding the parental characteristics for model 3 we appear to absorb part of the effects attributed to the young persons' characteristics, other school experiences, and family history in the first two specifications.

5.2 Obtaining an ATAR score

Table 4 presents average marginal effects of the middle-years experiences and other observable factors on the probability of obtaining and ATAR score, conditional on completing Year 12. As in Table 3, estimates for three different versions of the model are presented, with the estimates in column 3 being preferred for the same reasons as set out in the previous section. The discussion therefore focuses on these estimates in particular.

In contrast to Table 3, Table 4 suggests there is no statistically significant association between middle-years marijuana use or suspension/expulsion from school and the probability of obtaining an ATAR score (once we condition on Year 12 completion). Nevertheless, the point estimates remain negative, with young people who report using marijuana in the middle years and who complete Year 12 being 8.2 percentage points less likely to obtain an ATAR score. Young people who report having been suspended/expelled from school and who complete Year 12 are 6 percentage points less likely to obtain an ATAR score than those who do not report marijuana use or suspension/expulsion respectively.[33]

For marijuana use this is close to the unconditional gap of 6.4 percentage points reported in Table 2. For suspension/expulsion, the implication is that 10.7 percentage points of 16.7 percentage point gap shown in Table 2 can be explained by other observable differences between young people that do and do not report suspension or expulsion from school.

Few of the other observed factors have significant relationships with the probability of obtaining an ATAR score given Year 12 completion. The most significant determinant is family income-support receipt, where those with heavy family income-support receipt (see Section 3) are 14 percentage points less likely to obtain an ATAR score compared to respondents whose parents have no income-support receipt. Those with moderate (early) family income-support receipt are 11 percentage points less likely to obtain an ATAR score. Parental smoking is associated with a lower chance of obtaining an ATAR score. Father's highest level of education is positively associated with the probability of obtaining an ATAR score.

[33] Note that the effect of marijuana usage increases as more controls are added, even though the effect statistically insignificant in all specifications. The effect of suspension/expulsion is quite stable across specifications, yet its statistical significance decreases as more controls come at the expense of smaller sample size (resulting in higher standard errors).

5.3 ATAR score obtained

Table 5 presents OLS estimates of the association between middle-years experiences and other observable factors on the ATAR score, conditional on obtaining an ATAR score. As in Table 4, estimates for three different versions of the model are presented, with the estimates in column 3 again being preferred and forming the basis of the following discussion. The fourth column of Table 5 is for a slight variation of the model in column 3 that takes account of the bounded nature of the ATAR score – bounded between 30 and 100 – and is included as an additional robustness check.[34]

In no specification does marijuana usage have any statistically significant association with the ATAR score, although there is a stable point estimate suggesting a gap of around 3 points. This is smaller than the unconditional difference of 6 points shown by Table 2, suggesting that some of that gap is the result of observable differences in characteristics between the two groups.

In contrast, suspension/expulsion always has a statistically significant, negative association with the ATAR score amongst the respondents, which only weakens slightly as more controls are added. In the preferred specification (column 3), the ATAR scores of those who have been suspended/expelled from school are on average 5.7 points lower than those who have not, only slightly lower than the unconditional gap of 7.8 points shown by Table 2.

Of the other variables, repeating a school year, coming from a low-SES family (proxied by intensive Income Support receipt over the young person's childhood), and Indigenous status are all negatively associated with ATAR score. Being female, living in a metropolitan area, mother's age, having a non-English-speaking-background (NESB) migrant parent, and father's highest level of education attained are all positively associated with ATAR score.

5.4 Studying, working, or neither studying nor working at age 20

Table 6 shows the marginal effects on the probability of being in each of the three activity states for respondents at age 20 (studying, working, and neither studying nor working). Only results for the preferred specification including the parental controls from the parent questionnaire are presented in this case. Also note that the marginal effects for each particular variable sum to zero across the activity states. In other words, a factor that is associated with

[34] The results in columns 1-3 described above ignore the fact that ATAR scores below 30 are reported as 30. A tobit model can be used to account for this censoring in the data. The tobit marginal effects, which are reported in the last column of Table 4, are very similar to the OLS results because the censoring only affects 11 out of the 764 observations in the sample.

a 5 percentage point higher probability of studying must be associated with a 5 percentage point *lower* probability of *not* studying.

Marijuana use is associated with a lower probability of studying and a higher probability of working at age 20 (by 6 percentage points respectively), but neither marginal effect is statistically significant. The relevant unconditional associations from Table 2 show gaps of 17.3 percentage points for probability of studying and 8.4 percentage points for probability of working, with the implication being that most of the former and some of the latter are explained by differences in the observed characteristics of the two groups.

Suspension/expulsion is associated with a 19.6 percentage point lower probability of studying, a 13.4 percentage point higher probability of working, and a 6.2 percentage point higher probability of neither working nor studying at age 20, all of which are statistically significant. These are close to the relevant unconditional associations from Table 2, with the partial exception of the gap for neither working nor studying which falls from 11 percentage points to 6.2 percentage points when observable differences between those suspended/expelled and those not suspended/expelled are controlled for.

Of the other variables, repeating a school year and gender (being female) are positively associated with the probability of neither working nor studying. Mother's age, having a parent being born overseas and being from a non-English speaking background are negatively associated with the probability of neither working nor studying. Living in a metropolitan area is positively associated with studying and negatively associated with working and neither working nor studying. Father's education level is positively associated with studying and negatively associated with working.

5.5 Summary

In summary, this section of the report has shown that suspension/expulsion is strongly correlated with the probability of Year 12 completion, the ATAR score, and activity status at age 20 amongst YIF respondents. This is the case even when a host of observable differences between those that have been suspended/expelled and those that have not are controlled for. In contrast, marijuana usage amongst respondents is only significantly associated with Year 12 completion.

The extent to which we can we interpret these associations as *causal effects* of the middle-years experiences on the educational and labour market outcomes, however, depends on the extent to which *unobserved differences* between the relevant groups that are themselves

correlated with outcomes remain uncontrolled for. And while the stability of the estimates across different versions of the models presented in Tables 3-5 is encouraging – suggesting that these middle-years experiences are little correlated with *observed* family characteristics, family income-support history and parental characteristics – we cannot rule out that unobserved individual-idiosyncratic characteristics could be driving part or all of the remaining associations. The next section examines how likely this is using the method first put forward by Altonji et al. (2005).

6. Are These Relationships Causal?

In this section we assess the extent to which the associations between the middle-years experiences and later educational outcomes outlined previously can be interpreted as capturing causal relationships as opposed to unobserved differences between those that do and those that do not have the middle-years experience (marijuana use or suspension/expulsion). We do so using the approach suggested by Altonji et al. (2005). Technical details are located in the Appendix.

The first step in this process is to investigate *'how strongly correlated with the outcomes would the unobserved differences between groups have to be to explain away the whole association found in Tables 3-6'?* Table 7 presents the results. Each horizontal block corresponds to a particular outcome,[35] the rows within the blocks correspond to the middle-years experience in question. The columns report the key statistical associations under different assumptions about the strength of the correlation of the unobservables, with the assumed correlation getting stronger from left to right.

Table 7 clearly demonstrates that the effects of the middle-years experiences on all four outcomes are generally sensitive to the imposed degree of correlation between the unobserved factors and the outcome in question. For example, a weak correlation ($\rho = -0.1$) is sufficient for the negative association between marijuana usage on obtaining an ATAR score to become insignificant. The most persistent effect is found for suspension/expulsion on Year 12 completion, where the statistical significance of the association remains until we assume a moderate degree of correlation in unobservables ($\rho = -0.4$). This analysis suggests that the effect of suspension/expulsion on Year 12 completion is the most likely to be

[35] Note that Altonji et al.'s (2005) approach does not accommodate multi-categorical dependent variable. For this analysis we thus collapse the 'economic activity' into a binary variable (studying or not studying).

causal.[36] Overall, however, none of the estimated associations between the middle-years experiences and these outcomes stand up particularly well to plausible assumptions about the unobservables. This leads to the question whether any of these associations can be interpreted as causal relationships.

The second step in this process is to first assess the extent of selection on observables (this is what drives the contrast in statistical associations in Tables 3-6 compared to Table 2). Then, under the assumption that unobserved differences between groups are unlikely to be more important than observed differences between groups, an assessment can be made about whether the relevant coefficients would be robust to adding together the impact of observables and the likely maximum impact of unobservables.

Table 8 presents the results. Each column corresponds to one of the four outcomes of interest and the two horizontal blocks correspond to the two middle-years experiences. Within each horizontal block the critical row is the fifth row in each case, which shows the ratio of the estimated coefficient – accounting for observed differences but not unobserved differences between groups – to the maximum likely effect of the unobservables under the above assumption. Altonji et al. (2005) suggest that a ratio of estimate to bias of between 0 and 1 is indicative of an association between two variables that can plausibly be entirely explained by unobservable differences between the groups.[37]

Because the picture is clearer in the case of suspension/expulsion, it is the best place to start the discussion of Table 8. For all four outcomes the ratio of estimate to bias is between 0 and 1. This suggests that any associations between suspension/expulsion and these outcomes that remains after controlling for own characteristics, family characteristics, parental characteristics and family income-support history as in Tables 3-6 are unlikely to be causal, but more likely stem from other factors not accounted for in Tables 3-6. This is consistent with the results shown in Table 7.

For marijuana use the picture is more mixed. A similar conclusion can be drawn for the associations with Year 12 completion and with the probability of studying at age 20. In other words, any associations between marijuana use and these outcomes that remains after

[36] In a robustness check where the effect of each experience is analysed when the other experience is included but treated as exogenous (as in Tables 3-5), the effect of the experience is generally smaller and tends to lose significance even when a smaller ρ is assumed.

[37] In order to focus on each middle-years behaviour separately for the purposes of the Altonji et al. analysis discussed here, we simplify the relevant regression models by excluding the other middle-years behaviour variable in each case. The estimated coefficients and marginal effects in Table 8 are not identical to those in Tables 3-6 for this reason (see the Appendix for further discussion of this point).

controlling for own characteristics, family characteristics, parental characteristics and family income-support history as in Tables 3-6 are unlikely to be causal, but more likely stem from other factors not accounted for in Tables 3-6. Again this is consistent with the results shown in Table 7, and also broadly in line with Chatterji (2006a), who also uses Altonji et al.'s (2005) approach to show that the negative association between high-school alcohol use and educational attainment at age 26 is unlikely to indicate a causal relationship.

Table 8 suggests a different interpretation, however, for the associations between marijuana use and obtaining an ATAR score, and ATAR score obtained. In both cases the ratio of estimate to bias is *negative*, which reflects the ratio of negative coefficients to positive implied biases. The implication is that young people who report using marijuana in the middle years have other characteristics that actually *increase* the probability of obtaining an ATAR score and the ATAR score obtained.[38] Although this seems counter-intuitive at first, this is not the first study to suggest that middle-years marijuana users in Australia may have other characteristics that are associated with higher levels of educational achievement (see Van Ours and Williams, 2009).

Given this interpretation, the estimated coefficients on marijuana use in Tables 4 and 5, and in the first row of Table 8, will actually *underestimate* the causal effect of marijuana use on obtaining an ATAR score and the ATAR score obtained. In other words, middle-years marijuana use amongst YIF respondents *reduces* the probability of obtaining an ATAR score by at least 8.2 percentage points (using the estimate in Table 4) or 10.1 percentage points (using the estimate in Table 8). Similarly, middle-years marijuana use by YIF respondents *reduces* the ATAR score obtained by at least 2.7 points (using the estimate in Table 5) or 4.4

[38] Recall the argument from Section 4 that finding differences in *observed* characteristics between the two groups might be an indication that differences in *unobserved* characteristics also exist between the two groups. Relative to non-users, marijuana users do have some observed characteristics that are positively associated with educational outcomes, e.g. higher maternal education levels (see Table A3 in the Appendix). This suggests similarly advantageous differences in unobserved characteristics such as motivation levels or parental support are plausible.

points (using the estimate in Table 8).[39] So for these outcomes at least, a negative causal impact of marijuana use cannot be ruled out.[40]

7. Differential Impacts by SES

An objective of this report is to assess the extent to which the strength of the associations between the two middle-years experiences (marijuana use and suspension/expulsion) and the four outcomes (Year 12 completion, obtaining an ATAR score, ATAR score obtained, activity at age 20) vary by SES. Our conjecture is that associations may be stronger for young people from low-SES families than for young people from high-SES families. Given the report's findings so far, this conjecture implies that the negative impact of marijuana use on the probability of obtaining an ATAR score and on the ATAR score obtained may be stronger for young people from low-SES families than for young people from high-SES families. Similarly, although there is no overall evidence here for negative causal impacts from middle-years marijuana use on the other outcomes, or from suspension/expulsion on any of the four outcomes, it does not necessarily follow that there are no negative causal impacts of these behaviours on these outcomes for low-SES (or, for that matter, for high-SES) young people. This section examines this issue.

Following Cobb-Clark et al. (2012), our proxy for SES is the binary dummy for heavy family income-support receipt already included in the regression models discussed in Sections 5 and 6 (and defined in Section 3). To simplify, the two dummies for moderate income-support receipt are dropped so that the single income-support receipt dummy now distinguishes between those with heavy income-support receipt (low SES) and those with moderate or no income-support receipt (high SES).[41]

[39] The reason this is not picked up in Table 7 is that we have assumed that marijuana users have other unobserved characteristics that reduce the probability of obtaining an ATAR score and the score obtained in Table 7. In other words, we have assumed a negative correlation between the unobservables and the outcomes. If we were to assume a positive correlation between the unobservables and the outcomes – essentially extending Table 7 to the left – then the estimated coefficients on marijuana use for these two outcomes become larger in absolute magnitude (more negative) and increase in statistical significance.

[40] Note that once we discard unmatched questionnaires and once we condition on Year 12 completion (and for ATAR score obtained, on obtaining an ATAR score) there are very few remaining middle-years marijuana users in the YIF sample, Hhence the cautious phrasing of this particular conclusion.

[41] Splitting the sample into the four Income Support categories identified to this point is not possible because of sample size constraints. But it is possible to estimate the relevant models on a sub-sample that merges the two moderate Income Support categories, which we might interpret as medium SES. The estimated associations between marijuana use and outcomes for this medium-SES group generally fall between those for the high-SES and low-SES groups reported in Tables 9 and 10, as we might expect. The pattern of results is not so clear for suspension/expulsion, however, and there is some tentative evidence that associations with educational outcomes may be strongest for those in the medium-SES group.

As a first pass, each relevant sample is split into high-SES and low-SES sub-samples using the SES dummy. The regression models behind Tables 3-6 are then re-estimated to see if the associations between the middle-years experiences and later outcomes vary between the two SES-based groups. Results are presented in Tables 9 and 10.

Table 3 indicated that both marijuana usage and suspension/expulsion from school were negatively and statistically significantly associated with Year 12 completion across the whole sample (marginal effects of -0.1 and -0.18 respectively). When we split the sample, however, only the low-SES group show a statistically significant association between marijuana usage and Year 12 completion, with a larger magnitude of -0.26. This is larger than the marginal effects for all the other variables included in the model. For the high-SES group, the marginal effect is a statistically insignificant -0.05, less than one fifth of the low-SES estimate. Consistent with our conjecture, the suggestion is that middle-years marijuana usage may be more detrimental to chances of completing Year 12 for YIF respondents from low-SES families.

The marginal effects for suspension/expulsion from school are closer across the two groups, with both statistically significant and negative as in Table 3, but even so there is still a seven percentage point difference in the estimates for the two groups. It is notable that the difference between the two estimates runs counter to our conjecture that low-SES young people would be more impacted by suspension/expulsion than high-SES young people. Although the difference between these estimates is not statistically significant, suspension/expulsion is more strongly correlated with Year 12 completion for the higher SES group than for the lower SES group.

There are similar patterns by SES in the other associations. We find a stronger association between obtaining an ATAR score and marijuana use for low SES compared to high SES, with a 12 percentage point gap between the estimates. The association between studying at age 20 and marijuana use is stronger for low SES compared to high SES, with a 9 percentage point gap (see Table 10). In contrast, the association between obtaining an ATAR score and suspension/expulsion is stronger for high SES compared to low SES, with a 5 percentage point gap between the estimates. The association between studying at age 20 and suspension/expulsion is stronger for high SES compared to low SES, with a 4 percentage point gap. For ATAR score obtained, however, these patterns appear to be reversed, although this is the outcome with the smallest sample sizes – because ATAR score obtained is

conditioned on Year 12 completion and obtaining an ATAR score – and these estimates are particularly imprecise.

To test for the presence of differential effects more formally we then estimate extended versions of these models on the full sample, first adding an interaction term between low SES and marijuana use, and second adding an interaction term between low SES and suspension/expulsion. The statistical significance of these interactive terms tells us the degree to which we can be certain that these associations do vary by SES. Summary results (for the key parameters only) for Year 12 completion, obtaining an ATAR score, and the ATAR score itself, are presented in Table 11.

None of the interaction terms have marginal effects that are statistically significant from zero at standard levels. Taken at face value the implication is that our conjecture about differential effects by SES is not supported by the evidence. Bear in mind, however, that small sample sizes here lead to imprecision (high standard errors) in our estimates, so that even large point estimates show up as statistically insignificant. Table 9 also indicates that there are big differences in the point estimates between the two groups, particularly for marijuana use and Year 12 completion, but also for other associations.

To test sensitivity to this particular SES proxy, we also estimate two alternative sets of interaction models, first replacing the dummy variable for heavy income-support receipt with a dummy for any income-support receipt, and second replacing the dummy for heavy income-support receipt with a measure of the occupational ranking of the mother (the ANU4 scale).[42] Again, the marginal effects for the interactive variables are by and large statistically insignificant, but there continue to be differences in point estimates by SES.

In summary, given sample size constraints and the associated imprecision of the estimates it is only possible to draw tentative conclusions regarding the differential impacts of the two middle-years experiences on later outcomes by SES. But our conjecture about differential effects of middle-years marijuana use by SES is, in most cases, supported by the evidence presented here. In other words we cannot rule out the possibility that middle-years marijuana usage may be more detrimental to chances of completing Year 12, obtaining an ATAR score, or studying at age 20 for individuals from low-SES families than for individuals from high-SES families. There is no support for our conjecture about differential effects of

[42] The ANU4 scale is a continuous measure of SES developed at the Australian National University (for more details see Jones and McMillan (2001)). Here the ANU4 scale is calculated from the current or most recent occupation. Results are available from the authors on request.

suspension/expulsion. Rather, it is the educational outcomes of high-SES young people that appear to be more impacted by suspension/expulsion than outcomes for low-SES young people. There is even some (very tentative) evidence that educational outcomes may be *most* impacted by suspension/expulsion for medium-SES young people, although splitting the sample into more than two categories further exacerbates problems associated with small samples.

8. Conclusions

This study analyses data from the YIF survey to examine the associations between two middle-years experiences (whether or not the respondent had used marijuana by age 14, and whether or not the respondent had been suspended or expelled from school) and four educational and labour market outcomes measured at age 20 (Year 12 completion, obtaining an ATAR score, the ATAR score obtained, and whether studying, working or neither studying nor working). The extent to which these associations can be interpreted as indicating causal relationships from the middle-years experiences to the later outcomes is then assessed.

Comparing mean educational outcomes between those respondents that do and do not report the two middle-years experiences shows up some very strong associations. All of the associations with educational outcomes are negative. Respondents who report using marijuana by the age of 14 and respondents who report having been suspended or expelled from school have lower Year 12 completion rates than their counterparts (by 26.5 percentage points and 31.6 percentage points respectively). They also have lower probabilities of obtaining an ATAR score (by 6.4 percentage points and 16.7 percentage points respectively), lower ATAR scores when a score is obtained (by 6 points and 7.8 points respectively), and lower probabilities of studying at age 20 (by 17.3 percentage points and 21.9 percentage points respectively).

The strength of these associations tends to decrease when observable differences in the characteristics of those respondents that do/do not have the middle-years experiences and their families are controlled for in a multivariate regression framework. In other words, part of the gaps in outcomes can be explained by observable differences in individual and family characteristics between those respondents that do and do not report each middle-years experience. But there are still many strong associations, and all associations with educational outcomes remain negative. After controlling for observable differences in this way, young people who report using marijuana by the age of 14 and young people who report having

been suspended or expelled from school still have lower Year 12 completion rates than their counterparts (by 10 percentage points and 18.3 percentage points respectively), lower probabilities of obtaining an ATAR score (by 8.2 percentage points and 6 percentage points respectively), lower ATAR scores when a score is obtained (by 2.7 points and 5.7 points respectively), and lower probabilities of studying at age 20 (by 6.3 percentage points and 19.6 percentage points respectively).

In the same way that observable differences in characteristics between those that do/do not report the middle-years experiences can explain part of the overall gaps in outcomes, so too can unobservable differences in characteristics between the different groups (e.g. differences in attitudes to risk between the different groups of young people). Because such differences are unobservable we cannot control for them in a multivariate regression framework. But we can make informal (but informed) assessments of their likely importance.

Such assessments strongly suggest that none of the remaining associations between respondent suspension/expulsion and later outcomes are likely to reflect causal impacts of suspension/expulsion on the outcomes (i.e. all associations can be plausibly explained by differences in unobservable factors between those that do and those that do not report suspension/expulsion). For marijuana use the picture is more mixed. For Year 12 completion and studying at age 20 there is little evidence here to suggest causal impacts. But causal effects cannot be ruled out from middle-years marijuana use on the probability of obtaining and ATAR score and on the ATAR score obtained. The evidence presented here suggests that middle-years marijuana use amongst YIF respondents reduces the probability of obtaining an ATAR score by at least 8.2 percentage points, and the ATAR score obtained by at least 2.7 points.

Evidence is also presented of negative associations between family SES and the young person's educational outcomes for the YIF sample. Using family income-support history to proxy for SES, the report shows that low-SES young people (those with a record of intensive family income-support receipt) are 23.1 percentage points less likely to complete Year 12 compared to high-SES young people (those with a record of no family income-support receipt). Similarly, the probability of obtaining an ATAR score is 14 percentage points lower, the ATAR score obtained is 4.6 points lower, and the probability of studying at age 20 is 19.3 percentage points lower.

There is also some evidence presented here suggesting that associations between middle-years marijuana use and educational and labour market outcomes are stronger for low-SES individuals than for high-SES individuals, and that associations between suspension/expulsion and education and labour market outcomes are stronger for high-SES individuals than for low-SES individuals.

To the extent that these findings from the YIF study would generalise to the wider population of young people in Australia, the research suggests a number of policy implications.

First, even in the absence of causal effects, the strong associations found in the data between these middle-years experiences and educational outcomes at age 20 suggest that those who tried marijuana in the middle years and those who have been suspended or expelled from school may be important groups to target for interventions intended at improving young people's educational outcomes. Targeting interventions at those suspended/expelled is likely to be more practical, however, given that data on marijuana use are unlikely to be routinely available. This group is also larger, with around one-fifth of YIF respondents reporting having been suspended or expelled at least once. Further, those respondents who were suspended/expelled are disproportionately from low-SES families, so targeting interventions on this group could help improve the relative educational outcomes of low-SES young people.

Second, the evidence presented here suggesting no causal effect between suspension/expulsion and an individual's educational outcomes, on average, may be useful for school principals taking difficult decisions regarding the use of such sanctions as disciplinary measures of last resort.

Third, these findings suggest that interventions aimed at reducing the proportion of young people who use marijuana in the middle years may have positive knock-on effects on the proportion of Year 12 completers who obtain a university entrance score, and on the entrance score obtained.

Finally, although the YIF data are very rich in detail on both young people and their parents, and are probably the best data source for this purpose currently available in Australia,[43] there are two important data-related caveats here that are worth restating. First, the available sample can be quite small when using matched young person-parent records and analysing

[43] The Longitudinal Study of Australian Children will plausibly become a valuable data resource for examining these kinds of issues as the child cohort (currently aged 12/13 years) gets older. Waves are currently planned up until age 18/19 years.

behaviours (such as marijuana use) where not all of those in the sample answer the relevant questions in the questionnaire. Although the sample still appears to be reasonably representative of the wider cohort, one consequence of small sample sizes is that point estimates will be subject to larger margins of error, and therefore less likely to appear statistically significant. Second, although the YIF data allow us to explicitly identify marijuana use during the middle-years, they do not allow us to explicitly identify suspension/expulsion during the middle years. So although many of the suspension/expulsions analysed here are likely to have taken place during the middle years, the conclusions regarding suspension/expulsion are best interpreted as the effects of suspension/exclusion *not* restricted to the middle years.

References

Ai, C. and Norton, E. (2003). Interaction terms in logit and probit models. *Economics Letters*, 80(1):123–129.

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

Barón, J. D. and Cobb-Clark, D. (2010). Are young peoples educational outcomes linked to their sense of control? Discussion Paper 4907, Institute for the Study of Labor.

Blyth, D., Simmons, R., and Carlton-Ford, S. (1983). The adjustment of early adolescents to school transitions. *Journal of Early Adolescence*, 3(1-2):105–120.

Bray, J., Zarkin, G., Ringwalt, C., and Qi, J. (2000). The relationship between marijuana initiation and dropping out of high school. *Health Economics*, 9(1):9–18.

Breunig, R., Cobb-Clark, D., Gorgens, T., Ryan, C., and Sartbayeva, A. (2009). User's guide to the Youth in Focus data version 2.0. Youth in Focus Project Discussion Paper Series 8, Research School of Economics, Australian National University.

Brook, J., Balka, E., and Whiteman, M. (1999). The risks for late adolescence of early adolescent marijuana use. *American Journal of Public Health*, 89(10):1549–1554.

Brown, B. B. (1998). Learning contexts beyond the classroom: Extracurricular activities, community organizations, and peer groups. In Borman, K. and Schneider, B., editors, *The adolescent years: Social influences and educational challenges*, pages 109–141. The National Society for the Study of Education, Chicago.

Burgess, S., Gardiner, K., and Propper, C. (2001). Growing up: School, family and area influences on adolescents' later life chances. CASE papers 249, Centre for Analysis of Social Exclusion, London School of Economics.

Burgess, S. and Propper, C. (1998). Early health related behaviours and their impact on later life chances: Evidence from the US. CASE papers 6, Centre for Analysis of Social Exclusion, London School of Economics.

Byrnes, J. (2003). Cognitive development during adolescence. In Gerald, R. A. and Berzonsky, M. D., editors, *Blackwell Handbook of Adolescence*, pages 227–246. Blackwell Publishing, Malden, MA.

Chatterji, P. (2006a). Does alcohol use during high school affect educational attainment?: Evidence from the national education longitudinal study. *Economics of Education Review*, 25(5):482–497.

Chatterji, P. (2006b). Illicit drug use and educational attainment. *Health Economics*, 15(5):489–511.

Cobb-Clark, D., Ryan, C., and Sartbayeva, A. (2012). Taking chances: The effect of growing up on welfare on the risky behaviour of young people. *Scandinavian Journal of Economics*, forthcoming.

Cook, T., Deng, Y., and Morgano, E. (2007). Friendship influences during early adolescence: The special role of friends' grade point average. *Journal of Research on Adolescence*, 17(2):325–356.

Costenbader, V. and Markson, S. (1998). School suspension: A study with secondary school students. *Journal of School Psychology*, 36(1):59–82.

Council of Australian Governments (2011). 2011-12 Commonwealth Budget.

Eccles, J. S. (1999). The development of children ages 6 to 14. *The Future of Children*, 9(2):30–44.

Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, A., and MacIver, D. (1993). Development during adolescence: The impact of stage environment fit on young adolescents' experiences in schools and in families. *American Psychologist*, 48(2):90–101.

Fergusson, D. and Woodward, L. (2000). Educational, psychosocial, and sexual outcomes of girls with conduct problems in early adolescence. *Journal of Child Psychology and Psychiatry*, 41(6):779–792.

French, M. and Popovici, I. (2011). That instrument is lousy! In search of agreement when using instrumental variables estimation in substance use research. *Health Economics*, 20(2):127–146.

Gruber, J. (2000). *Risky behaviour among youths: An economic analysis*. University of Chicago Press, Chicago.

Heckman, J. (2000). Policies to foster human capital. *Research in Economics*, 54(1):3–56.

Heckman, J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, 312(5782):1900–1902.

Homel, J., Mavisakalyan, A., Nguyen, H.T. and Ryan, C. (2012). School completion: What we learn from different measures of family background. NCVER, Adelaide.

Israel, M. and Seeborg, M. (1998). The impact of youth characteristics and experiences on transitions out of poverty. *Journal of Socio-Economics*, 27(6):753–776.

Jones, F. and McMillan, J. (2001). Scoring occupational categories for social research: A review of current practice, with Australian examples. *Work, Employment & Society*, 15(3):539–563.

Juvonen, J., Le, V., Kaganoff, T., Augustine, C., and Constant, L. (2004). *Focus on the wonder years: Challenges facing the American middle school*, volume 139. Rand Corporation.

Kaplan, D., Liu, R., and Kaplan, H. (2005). School related stress in early adolescence and academic performance three years later: The conditional influence of self expectations. *Social Psychology of Education*, 8(1):3–17.

Karaca-Mandic, P., Norton, E., and Dowd, B. (2012). Interaction terms in nonlinear models. *Health Services Research*, 47(1):255–275.

Karakus, M. C., Salkever, D. S., Slade, E. P., Ialongo, N., and Stuart, E. (2012). Implications of middle school behaviour problems for high school graduation and employment outcomes of young adults: Estimation of a recursive model. *Education Economics*, 20(1):33–52.

Kasen, S., Cohen, P., and Brook, J. (1998). Adolescent school experiences and dropout, adolescent pregnancy, and young adult deviant behaviour. *Journal of Adolescent Research*, 13(1):49–72.

Lenzi, M., Vieno, A., Perkins, D., Pastore, M., Santinello, M., and Mazzardis, S. (2011). Perceived neighbourhood social resources as determinants of pro-social behaviour in early adolescence. *American Journal of Community Psychology*, pages 1–13.

Mendez, L. (2003). Predictors of suspension and negative school outcomes: A longitudinal investigation. *New directions for youth development*, 2003(99):17–33.

Mendez, R. and Sanders, S. (1981). An examination of in-school suspension: Panacea or Pandora's box? *NASSP Bulletin*, 65(441):65–69.

Morrison, G., Anthony, S., Storino, M., Cheng, J., Furlong, M., and Morrison, R. (2001). School expulsion as a process and an event: Before and after effects on children at risk for school discipline. *New Directions for Mental Health Services,* 2001(92):45–71.

Norton, E., Wang, H., and Ai, C. (2004). Computing interaction effects and standard errors in logit and probit models. *Stata Journal,* 4:154–167.

Patton, G.C., Coffey, C., Carlin, J.B., Degenhardt, L., Lynskey, M. and Hall, W. (2002). Cannabis use and mental health in young people: cohort study. *British Medical Journal*, 325 (7374):1195–1198.

Register, C.A., Williams, D.R. and Grimes, P.W. (2001). Adolescent drug use and educational attainment. Education Economics, 9(1), 1–18.

Renda, J., Vassallo, S., and Edwards, B. (2011). Bullying in early adolescence and its association with anti-social behaviour, criminality and violence 6 and 10 years later. *Criminal Behaviour and Mental Health,* 21(2):117–127.

Sandefur, G., McLanahan, S., and Wojtkiewicz, R. (1992). The effects of parental marital status during adolescence on high school graduation. *Social Forces,* 71(1):103–121.

Scheier, L. and Botvin, G. (1995). Effects of early adolescent drug use on cognitive efficacy in early-late adolescence: A developmental structural model. *Journal of Substance Abuse,* 7(4):379–404.

Smetana, J. and Turiel, E. (2003). Moral development during adolescence. In Gerald, R. A. and Berzonsky, M. D., editors, *Blackwell Handbook of Adolescence,* pages 247–268. Blackwell Publishing, Malden, MA.

Tomlinson, M. and Walker, R. (2010). Poverty, adolescent well-being and outcomes later in life. *Journal of International Development,* 22(8):1162–1182.

Wigfield, A., Eccles, J., Roeser, R., and Schiefele, U. (2006). Development of achievement motivation. In Damon, W., Lerner, R. M., and Eisenberg, N., editors, H*andbook of Child Psychology,* pages 406–434. John Wiley, New York, 6th edition.

Van Ours, J. and Williams, J. (2009). Why parents worry: Initiation into cannabis use by youth and their educational attainment. *Journal of Health Economics,* 28(1):132–142.

Tables

Table 1: Sample means for outcomes, middle-years experiences, and other characteristics

	Total sample (weighted)	Total sample (unweighted)	With matched parental records (unweighted)
Outcomes			
Completed Year 12	0.792	0.772	0.800***
Obtained an ATAR score	0.812	0.8	0.812
ATAR score (scale: 0-100)	75.001	74.533	75.260
Studying	0.582	0.562	0.579**
Working	0.33	0.338	0.331
Neither studying nor working	0.088	0.1	0.090**
Middle-years experiences			
Tried marijuana by age 14	0.078	0.084	0.086
Suspended/expelled from school	0.196	0.204	0.187**
Other experiences			
Repeated a year	0.105	0.112	0.106
Number of schools	3.026	3.102	2.955***
Own characteristics			
Female	0.49	0.554	0.553
Indigenous Australian	0.033	0.038	0.034
Born in a NESB country	0.077	0.082	0.042***
Metropolitan residence	0.666	0.668	0.629***
Family characteristics at age 14			
Lived with both parents at 14	0.738	0.686	0.732***
Mother employed at 14	0.703	0.672	0.708***
Family income-support history			
Heavy family income-support receipt	0.275	0.353	0.305***
Moderate (early) family income-support receipt	0.234	0.247	0.259
Moderate (late) family income-support receipt	0.081	0.129	0.123
Parental characteristics			
Age of mother	47.039		46.802
Number of children of mother	2.951		3.004
Mother is a smoker	0.188		0.214
At least one parent born in a NESB country	0.244	0.256	0.190***
Mother's highest education: Year 12 or equiv.	0.086		0.085
Mother's highest education: above Year 12	0.654		0.641
Father's highest education: Year 12 or equiv.	0.149		0.149
Father's highest education: above Year 12	0.461		0.439
Number of observations	3,560	3,560	1,819

Notes: Parental migrant status (born overseas) is reported by the young person. All other parental characteristics are reported by the parent. *, ** and *** denote sample means that are significantly different from the total sample at the 10%, 5% and 1% level respectively.

Table 2: Outcomes by experiences and family characteristics

	Number of obs.	Compl. Year 12	Obtained ATAR	ATAR score	Studying	Working	Neither studying nor working
Middle-years experiences							
Tried marijuana by age 14							
0. No	2,235	87	85.4	76.7	64.8	29	6.2
1. Yes	204	60.5***	79.0*	70.7***	47.5***	37.4	15.1
Suspended/expelled from school							
0. No	2,829	85.4	83.3	75.8	62.5	30.9	6.7
1. Yes	725	53.8***	66.6***	68.0***	40.6***	41.7	17.7
Other experiences							
Repeated a year							
0. No	3,155	81.8	82.5	75.2	59.5	32.8	7.7
1. Yes	399	56.8***	62.7***	71.0**	46.1***	35.6	18.3
Number of schools							
1. 0-3 schools	2,589	82.5	82.5	74.8	60.8	31.9	7.4
2. 4 schools	463	73.2***	76.4**	76.6	48.6***	39.4	11.9
3. 5+ schools	500	66.2***	75.7***	74.7	51.8***	34.1	14.1
Family characteristics at age 14							
Lived with both parents at 14							
0. No	1,115	65.2	74.2	72.2	47.4	38	14.5
1. Yes	2,437	84.1***	83.1***	75.7***	61.9***	31.3	6.8
Mother employed at 14							
0. No	1,116	75.2	78.5	75.2	55.9	32.7	11.4
1. Yes	2,284	82.1***	82.6**	75.0	59.5*	33.2	7.3
Family income-support history							
1. No income-support receipt	965	88.9	86.8	76.5	67.2	28.3	4.5
2. Intensive receipt	1,256	65.8***	72.8***	71.9***	47.9***	36.6	15.5
3. Moderate (late) receipt	458	77.5***	77.8***	74.5*	54.0***	36.9	9.1
4. Moderate (early) receipt	881	80.5***	81.2*	74.2	59.2**	33.6	7.2

Notes: Entries are percentages, except for mean of ATAR score. All statistics are weighted. *, ** and *** denote statistics that are significantly different from the relevant reference category (i.e. the first category for each variable) at the 10%, 5% and 1% level respectively.

Table 3: Marginal effects of middle-years experiences and other factors on the probability of completing Year 12

	(1)	(2)	(3)	(4)
Middle-years experiences				
Tried marijuana by age 14	-0.125***	-0.115***	-0.100**	-0.098**
	(0.031)	(0.031)	(0.043)	(0.043)
Suspended/expelled from school	-0.213***	-0.180***	-0.183***	-0.210***
	(0.026)	(0.026)	(0.038)	(0.039)
Other experiences				
Repeated a year	-0.150***	-0.150***	-0.131***	-0.151***
	(0.029)	(0.030)	(0.042)	(0.044)
Number of schools	-0.020***	-0.011***	-0.010*	-0.010*
	(0.004)	(0.004)	(0.006)	(0.006)
Own characteristics				
Female	0.012	0.022	0.027	0.019
	(0.014)	(0.015)	(0.019)	(0.019)
Indigenous Australian	-0.059	-0.029	0.033	0.011
	(0.046)	(0.042)	(0.043)	(0.051)
Born in a NESB country	0.081***	0.085***	0.024	0.074**
	(0.022)	(0.021)	(0.053)	(0.033)
Metropolitan residence	0.067***	0.064***	0.032*	0.049**
	(0.016)	(0.016)	(0.019)	(0.020)
Family characteristics at age 14				
Lived with both parents at 14		0.049***	-0.006	-0.006
		(0.018)	(0.023)	(0.023)
Mother employed at 14		0.039**	0.021	0.038*
		(0.016)	(0.021)	(0.022)
Family income-support history				
Heavy family income-support receipt		-0.107***	-0.072**	-0.108***
		(0.024	(0.031)	(0.033)
Moderate (early) family income-support receipt		-0.070***	-0.034	-0.055*
		(0.023)	(0.027)	(0.028)
Moderate (late) family income-support receipt		-0.057*	-0.020	-0.037
		(0.029)	(0.034)	(0.037)
Parental characteristics				
Age of mother			0.002	
			(0.002)	
Number of children of mother			-0.015**	
			(0.007)	
Mother is a smoker			-0.034	
			(0.025)	
At least one parent born in a NESB country			0.039*	
			(0.023)	
Mother's highest education: Year 12 or equiv.			0.035	
			(0.028)	
Mother's highest education: above Year 12			0.041*	
			(0.022)	
Father's highest education: Year 12 or equiv.			0.005	
			(0.026)	
Father's highest education: above Year 12			0.050**	
			(0.020)	
Observations	2419	2330	1182	1182
Pseudo R^2	0.149	0.177	0.209	0.180

Notes: Standard errors are in parentheses. *significant at 10%, **significant at 5%, ***significant at 1%.

Table 4: Marginal effects of middle-years experiences and other factors on the probability of obtaining an ATAR score

	(1)	(2)	(3)
Middle-years experiences			
Tried marijuana by age 14	-0.029	-0.049	-0.082
	(0.038)	(0.041)	(0.059)
Suspended/expelled from school	-0.078**	-0.057*	-0.060
	(0.032)	(0.031)	(0.044)
Other experiences			
Repeated a year	-0.172***	-0.147***	-0.050
	(0.044)	(0.044)	(0.054)
Number of schools	-0.011*	-0.007	-0.012
	(0.006)	(0.006)	(0.008)
Own characteristics			
Female	0.013	0.022	-0.002
	(0.018)	(0.018)	(0.023)
Indigenous Australian	-0.054	-0.036	0.015
	(0.064)	(0.061)	(0.069)
Born in a NESB country	0.081***	0.085***	0.029
	(0.024)	(0.023)	(0.057)
Metropolitan residence	0.059***	0.061***	0.023
	(0.020)	(0.020)	(0.024)
Family characteristics at age 14			
Lived with both parents at 14		0.010	-0.041
		(0.021)	(0.026)
Mother employed at 14		0.030	-0.005
		(0.019)	(0.026)
Family income-support history			
Heavy family income-support receipt		-0.136***	-0.137***
		(0.029)	(0.042)
Moderate (early) family income-support receipt		-0.081***	-0.113***
		(0.027)	(0.035)
Moderate (late) family income-support receipt		-0.064*	-0.033
		(0.034)	(0.043)
Parental characteristics			
Age of mother			0.002
			(0.002)
Number of children of mother			-0.008
			(0.009)
Mother is a smoker			-0.073**
			(0.034)
At least one parent born in a NESB country			0.037
			(0.029)
Mother's highest education: Year 12 or equiv.			0.038
			(0.036)
Mother's highest education: above Year 12			0.019
			(0.027)
Father's highest education: Year 12 or equiv.			0.027
			(0.030)
Father's highest education: above Year 12			0.069***
			(0.026)
Observations	1872	1818	960
R^2			
Pseudo R^2	0.038	0.062	0.087

Notes: Standard errors are in parentheses. *significant at 10%, **significant at 5%, ***significant at 1%.

Table 5: Marginal effects of middle-years experiences and other factors on the ATAR score obtained

	(1)	(2)	(3)	(4)
Middle-years experiences				
Tried marijuana by age 14	-3.288	-3.181	-2.722	-2.716
	(2.001)	(2.062)	(2.811)	(2.764)
Suspended/expelled from school	-6.633***	-6.024***	-5.662*	-5.649*
	(1.552)	(1.576)	(2.306)	(2.264)
Other experiences				
Repeated a year	-5.271*	-5.377*	-6.643*	-6.624*
	(2.131)	(2.120)	(2.835)	(2.781)
Number of schools	0.446	0.669*	0.089	0.088
	(0.318)	(0.325)	(0.490)	(0.482)
Own characteristics				
Female	1.407	1.463	3.099**	3.094**
	(0.885)	(0.892)	(1.167)	(1.148)
Indigenous Australian	-8.386*	-8.888*	-4.213	-4.203
	(3.827)	(3.927)	(4.734)	(4.650)
Born in a NESB country	0.410	0.648	2.171	2.169
	(1.434)	(1.462)	(2.920)	(2.875)
Metropolitan residence	1.472	1.530	2.028	2.025
	(0.983)	(0.992)	(1.233)	(1.213)
Family characteristics at age 14				
Lived with both parents at 14		0.223	-0.233	-0.233
		(1.157)	(1.623)	(1.597)
Mother employed at 14		-0.097	-0.305	-0.305
		(0.983)	(1.345)	(1.324)
Family income-support history				
Heavy family income-support receipt		-4.958***	-3.200	-3.195
		(1.243)	(1.782)	(1.752)
Moderate (early) family income-support receipt		-2.528*	-2.032	-2.028
		(1.141)	(1.491)	(1.467)
Moderate (late) family income-support receipt		-2.605	-1.298	-1.296
		(1.363)	(1.733)	(1.704)
Parental characteristics				
Age of mother			0.244	0.244*
			(0.125)	(0.123)
Number of children of mother			-0.063	-0.063
			(0.493)	(0.485)
Mother is a smoker			-2.185	-2.181
			(1.729)	(1.700)
At least one parent born in a NESB country			3.092*	3.089*
			(1.570)	(1.545)
Mother's highest education: Year 12 or equiv.			3.835	3.830
			(2.187)	(2.154)
Mother's highest education: above Year 12			1.838	1.835
			(1.449)	(1.425)
Father's highest education: Year 12 or equiv.			0.288	0.288
			(1.795)	(1.767)
Father's highest education: above Year 12			4.450***	4.443***
			(1.326)	(1.305)
Observations	1435	1399	764	764
R^2	0.030	0.044	0.096	
Pseudo R^2				0.012

Notes: Specification (4) is estimated by a tobit model, other specifications are estimated by OLS. Standard errors are in parentheses. *significant at 10%, **significant at 5%, ***significant at 1%.

Table 6: Marginal effects of middle-years experiences and other factors on the probability of studying, working, or neither working nor studying

	Studying	Working	Neither studying nor working
	(1)	(2)	(3)
Middle-years experiences			
Tried marijuana by age 14	-0.063	0.062	0.000
	(0.057)	(0.055)	(0.022)
Suspended/expelled from school	-0.196***	0.134***	0.062**
	(0.046)	(0.045)	(0.027)
Other experiences			
Repeated a year	-0.037	-0.038	0.075**
	(0.052)	(0.048)	(0.032)
Number of schools	0.004	-0.003	-0.001
	(0.010)	(0.010)	(0.004)
Own characteristics			
Female	-0.016	-0.022	0.038**
	(0.028)	(0.027)	(0.013)
Indigenous Australian	-0.060	0.082	-0.022
	(0.091)	(0.090)	(0.022)
Born in a NESB country	-0.069	-0.083	0.151
	(0.089)	(0.071)	(0.093)
Metropolitan residence	0.080***	-0.052*	-0.028*
	(0.029)	(0.028)	(0.014)
Family characteristics at age 14			
Lived with both parents at 14	0.027	-0.016	-0.012
	(0.037)	(0.036)	(0.017)
Mother employed at 14	0.031	-0.010	-0.021
	(0.032)	(0.031)	(0.016)
Family income-support history			
Heavy family income-support receipt	-0.015	0.005	0.010
	(0.041)	(0.040)	(0.022)
Moderate (early) family income-support receipt	-0.054	0.019	0.035
	(0.037)	(0.036)	(0.024)
Moderate (late) family income-support receipt	-0.037	0.032	0.005
	(0.047)	(0.046)	(0.027)
Parental characteristics			
Age of mother	0.001	0.002	-0.003**
	(0.003)	(0.003)	(0.001)
Number of children of mother	0.000	-0.007	0.006
	(0.011)	(0.011)	(0.005)
Mother is a smoker	-0.018	-0.016	0.034*
	(0.037)	(0.035)	(0.019)
At least one parent born in a NESB country	0.044	-0.012	-0.032*
	(0.038)	(0.037)	(0.018)
Mother's highest education: Year 12 or equiv.	-0.020	0.047	-0.027
	(0.051)	(0.050)	(0.019)
Mother's highest education: above Year 12	0.053	-0.054	0.001
	(0.034)	(0.033)	(0.014)
Father's highest education: Year 12 or equiv.	0.008	-0.002	-0.006
	(0.040)	(0.039)	(0.019)
Father's highest education: above Year 12	0.092***	-0.072**	-0.020
	(0.032)	(0.031)	(0.015)

Notes: N=1,181. Pseudo R-squared = 0.074. Standard errors are in parentheses. *significant at 10%, **significant at 5%, ***significant at 1%.

Table 7: Effect of middle-years experience on outcome under varying degrees of correlation between the unobserved determinants of the two

	ρ =0	ρ =-0.1	ρ =-0.2	ρ =-0.3	ρ =-0.4	ρ =-0.5
Year 12 completion						
Marijuana use	-0.804***	-0.609***	-0.414***	-0.217	-0.018	0.181
	(0.160)	(0.160)	(0.158)	(0.155)	(0.151)	(0.145)
	[-0.197]	[-0.140]	[-0.089]	[-0.044]	[-0.003]	[0.032]
Suspension/ Expulsion	-0.789***	-0.611***	-0.430***	-0.248***	-0.063	0.123
	(0.095)	(0.095)	(0.094)	(0.092)	(0.089)	(0.086)
	[-0.214]	[-0.159]	[-0.108]	[-0.060]	[-0.015]	[0.027]
Obtaining an ATAR score						
Marijuana use	-0.410*	-0.206	-0.005	0.193	0.387*	0.578***
	(0.216)	(0.215)	(0.213)	(0.209)	(0.204)	(0.196)
	[-0.101]	[-0.047]	[-0.001]	[0.037]	[0.069]	[0.094]
Suspension/Expulsion	-0.348***	-0.165	0.017	0.198	0.377***	0.554***
	(0.130)	(0.129)	(0.128)	(0.125)	(0.122)	(0.117)
	[-0.092]	[-0.041]	[0.004]	[0.044]	[0.080]	[0.112]
ATAR score						
Marijuana use	-4.385	-1.378	1.659	4.761*	7.978***	11.384***
	(2.700)	(2.70)	(2.680)	(2.654)	(2.617)	(2.568)
Suspension/Expulsion	-8.119***	-5.275***	-2.393	0.570	3.671**	6.992***
	(1.831)	(1.829)	(1.823)	(1.814)	(1.800)	(1.782)
Studying at age 20						
Marijuana use	-0.375**	-0.181	0.014	0.209	0.405***	0.600***
	(0.146)	(0.145)	(0.144)	(0.141)	(0.137)	(0.132)
	[-0.139]	[-0.066]	[0.005]	[0.071]	[0.132]	[0.186]
Suspension/expulsion	-0.548***	-0.369***	-0.187**	-0.003	0.183**	0.371***
	(0.089)	(0.089)	(0.088)	(0.086)	(0.084)	(0.081)
	[-0.208]	[-0.139]	[-0.070]	[-0.001]	[0.066]	[0.131]

Notes: Standard errors are in parentheses. Marginal effects in square brackets. *significant at 10%, **significant at 5%, ***significant at 1%.

Table 8: Amount of selection on unobservables relative to selection on observables required to attribute the entire effect of middle-years experiences to selection bias

	Estimates	Year 12 completion	Obtaining an ATAR score	ATAR score	Studying at age 20
Marijuana use	Unconstrained estimate	-0.804***	-0.410*	-4.385	-0.375**
	Standard error	(0.160)	(0.216)	(2.738)	(0.146)
	Marginal effect	[-0.197]	[-0.101]		[-0.139]
	Implied bias	-1.110	0.571	15.602	-1.196
	Ratio of estimate to bias	0.724	-0.718	-0.281	0.314
	Sample size	1,182	960	764	1,181
Suspension/Expulsion	Unconstrained estimate	-0.789***	-0.348***	-8.119***	-0.548***
	Standard error	(0.095)	(0.130)	(1.851)	(0.089)
	Marginal effect	[-0.214]	[-0.092]		[-0.208]
	Implied bias	-1.597	-1.296	-19.030	-2.290
	Ratio of estimate to bias	0.494	0.269	0.427	0.239
	Sample size	1,643	1,255	956	1,642

Notes: Standard errors are in parentheses. Marginal effects in square brackets. *significant at 10%, **significant at 5%, ***significant at 1%.

Table 9: (Marginal) effects of middle-years experiences on education outcomes by SES

	Year 12 completion		Obtaining ATAR score		ATAR score	
	High SES	Low SES	High SES	Low SES	High SES	Low SES
	(1)	(2)	(3)	(4)	(5)	(6)
Middle-years experiences						
Tried marijuana by age 14	-0.048	-0.257**	-0.025	-0.142	-1.782	-1.364
	(0.043)	(0.106)	(0.055)	(0.159)	(3.014)	(9.322)
Suspended/expelled from school	-0.216***	-0.140**	-0.079	-0.026	-5.488	-7.353
	(0.048)	(0.071)	(0.056)	(0.083)	(2.826)	(4.264)
Other experiences						
Repeated a year	-0.147***	-0.149*	-0.059	-0.012	-7.204*	-7.087
	(0.054)	(0.081)	(0.065)	(0.114)	(3.407)	(5.456)
Number of schools	-0.004	-0.020	-0.010	-0.027	0.015	0.147
	(0.007)	(0.013)	(0.009)	(0.019)	(0.568)	(1.027)
Own characteristics						
Female	0.025	0.039	0.012	-0.032	3.875**	0.959
	(0.020)	(0.046)	(0.025)	(0.059)	(1.308)	(2.801)
Indigenous Australian	0.022	0.028	-0.011	0.050	-5.503	0.420
	(0.061)	(0.088)	(0.096)	(0.134)	(5.578)	(9.306)
Born in a NESB country			0.012	0.069	2.038	1.011
			(0.069)	(0.132)	(3.602)	(5.445)
Metropolitan residence	0.004	0.097**	0.002	0.077	2.042	1.572
	(0.020)	(0.046)	(0.025)	(0.060)	(1.373)	(2.941)
Family characteristics at age 14						
Lived with both parents at 14	-0.025	0.007	0.017	-0.151**	-0.119	-0.734
	(0.025)	(0.047)	(0.037)	(0.059)	(2.015)	(2.968)
Mother employed at 14	0.022	0.014	0.017	-0.061	-0.883	1.566
	(0.023)	(0.046)	(0.030)	(0.059)	(1.553)	(2.961)
Parental characteristics						
Age of mother	0.003	0.001	0.002	0.004	0.154	0.491*
	(0.002)	(0.004)	(0.003)	(0.005)	(0.149)	(0.239)
Number of children of mother	-0.025**	0.010	-0.006	-0.014	-0.296	0.349
	(0.007)	(0.016)	(0.011)	(0.017)	(0.601)	(0.881)
Mother is a smoker	-0.022	-0.082	-0.114***	-0.001	-4.122*	1.814
	(0.028)	(0.055)	(0.044)	(0.067)	(2.083)	(3.222)
At least one parent born in a NESB country	0.036	0.055	0.023	0.134**	2.834	3.035
	(0.023)	(0.054)	(0.032)	(0.063)	(1.800)	(3.488)
Mother's highest edu: Year 12	0.018	0.093	0.007	0.140*	5.205*	-3.461
	(0.029)	(0.066)	(0.043)	(0.072)	(2.498)	(4.882)
Mother's highest edu.: above Yr 12	0.052**	0.021	0.002	0.103	3.481*	-5.224
	(0.024)	(0.049)	(0.029)	(0.066)	(1.637)	(3.314)
Father's highest edu: Year 12	0.022	-0.051	0.023	0.030	0.381	-0.791
	(0.024)	(0.068)	(0.032)	(0.081)	(2.057)	(3.849)
Father's highest edu.: above Yr 12	0.052**	0.057	0.053*	0.108*	3.615*	8.188**
	(0.022)	(0.049)	(0.029)	(0.059)	(1.504)	(2.978)
Observations	872	310	742	218	614	150
R^2					0.091	0.155
Pseudo R^2	0.212	0.202	0.049	0.136		

Notes: Standard errors are in parentheses. *significant at 10%, **significant at 5%, ***significant at 1%. High SES: no to moderate family income-support receipt; Low SES: intensive family income-support receipt.

Table 10: Marginal effects of middle-years experiences on probability of studying, working, or neither working nor studying, by SES

	Studying		Working		Neither studying nor working	
	High SES	Low SES	High SES	Low SES	High SES	Low SES
	(1)	(2)	(3)	(4)	(5)	(6)
Middle-years experiences						
Tried marijuana by age 14	-0.037	-0.128	0.058	0.054	-0.021	0.075
	(0.067)	(0.106)	(0.066)	(0.102)	(0.019)	(0.067)
Suspended/expelled from school	-0.216***	-0.173*	0.103*	0.151*	0.113***	0.022
	(0.057)	(0.079)	(0.054)	(0.080)	(0.042)	(0.042)
Other experiences						
Repeated a year	-0.079	0.041	0.019	-0.141**	0.060	0.100*
	(0.067)	(0.082)	(0.063)	(0.071)	(0.038)	(0.059)
Number of schools	0.004	0.005	-0.000	-0.007	-0.004	0.002
	(0.013)	(0.017)	(0.013)	(0.016)	(0.005)	(0.008)
Own characteristics						
Female	-0.003	-0.059	-0.027	-0.015	0.030**	0.074***
	(0.032)	(0.056)	(0.032)	(0.055)	(0.014)	(0.028)
Indigenous Australian	0.186*	-0.314**	-0.139	0.313**	-0.047***	0.000
	(0.095)	(0.138)	(0.095)	(0.139)	(0.007)	(0.050)
Born in a NESB country	-0.178	0.042	-0.131*	0.043	0.309*	-0.086***
	(0.134)	(0.145)	(0.076)	(0.145)	(0.158)	(0.015)
Metropolitan residence	0.073**	0.076	-0.035	-0.092*	-0.039**	0.016
	(0.034)	(0.057)	(0.033)	(0.056)	(0.016)	(0.029)
Family characteristics at age 14						
Lived with both parents at 14	0.076	-0.064	-0.044	0.040	-0.032	0.024
	(0.050)	(0.057)	(0.049)	(0.056)	(0.026)	(0.034)
Mother employed at 14	0.028	0.012	-0.009	-0.006	-0.019	-0.006
	(0.038)	(0.057)	(0.037)	(0.056)	(0.018)	(0.033)
Parental characteristics						
Age of mother	0.000	0.003	0.002	0.003	-0.003	-0.006*
	(0.004)	(0.005)	(0.003)	(0.005)	(0.002)	(0.003)
Number of children of mother	-0.012	0.027	0.003	-0.021	0.010*	-0.006
	(0.013)	(0.019)	(0.013)	(0.019)	(0.005)	(0.011)
Mother is a smoker	0.020	-0.117*	-0.038	0.033	0.018	0.084**
	(0.045)	(0.064)	(0.042)	(0.063)	(0.022)	(0.037)
At least one parent born in a NESB country	0.038	0.068	-0.002	-0.035	-0.037**	-0.032
	(0.045)	(0.075)	(0.043)	(0.073)	(0.018)	(0.040)
Mother's highest education: Year 12 or equiv.	-0.020	-0.001	0.054	-0.000	-0.035**	0.001
	(0.059)	(0.097)	(0.058)	(0.095)	(0.016)	(0.057)
Mother's highest education: above Year 12	0.059	0.054	-0.062	-0.030	0.003	-0.024
	(0.040)	(0.063)	(0.039)	(0.061)	(0.016)	(0.033)
Father's highest education: Year 12 or equiv.	0.017	-0.018	-0.023	0.047	0.006	-0.029
	(0.046)	(0.079)	(0.043)	(0.079)	(0.022)	(0.041)
Father's highest education: above Year 12	0.092**	0.099	-0.091**	-0.036	-0.001	-0.062**
	(0.037)	(0.063)	(0.036)	(0.062)	(0.017)	(0.029)
Observations	871	310	871	310	871	310
Pseudo R^2	0.074	0.138	0.074	0.138	0.074	0.138

Notes: Standard errors are in parentheses. *significant at 10%, **significant at 5%, ***significant at 1%. High SES: no to moderate family income-support receipt; Low SES: intensive family income-support receipt.

Table 11: Interaction effects of middle-years experiences on education outcomes by SES

	Year 12 completion	Obtaining an ATAR score	ATAR score
Marijuana use			
Tried marijuana by age 14 (high SES)	-0.06	-0.05	-2.06
	(0.05)	(0.06)	(2.94)
Low SES (did not try marijuana before 15)	-0.04*	-0.07**	-2.06
	(0.02)	(0.03)	(1.63)
Tried marijuana by age 14 × low SES	-0.11	-0.07	-4.33
	(0.09)	(0.16)	(9.46)
Suspension/expulsion			
Ever suspended/expelled (high SES)	-0.21***	-0.08	-5.67**
	(0.05)	(0.06)	(2.79)
Low SES (never suspended/excluded)	-0.06**	-0.07**	-2.14
	(0.03)	(0.03)	(1.68)
Ever suspended/expelled × low SES	0.05	0.02	-0.23
	(0.07)	(0.10)	(4.85)
Observations	1182	960	764
Pseudo R^2	0.21	0.07	0.09

Notes: Standard errors are in parentheses. *significant at 10%, **significant at 5%, ***significant at 1%. High SES: no to moderate family income-support receipt; Low SES: intensive family income-support receipt.

Appendix: Additional Statistics and Methodological Details

A.1 Additional statistics

Table A1: Sample means by economic activity

	Studying and working	Studying only	Working	Neither studying nor working
Outcomes				
Completed Year 12	0.884	0.863***	0.688***	0.459***
Obtained an ATAR score	0.870	0.838**	0.699***	0.607***
ATAR score (scale: 0-100)	77.234	77.898	67.318***	66.765***
Middle-years experiences				
Tried marijuana by age 14	0.068	0.037***	0.101**	0.194***
Suspended/expelled from school	0.128	0.169**	0.256***	0.402***
Other experiences				
Repeated a year	0.076	0.124	0.119	0.230***
Number of schools	2.938	3.228*	3.121	3.599***
Own characteristics				
Female	0.554	0.560	0.533*	0.612**
Indigenous Australian	0.017	0.046	0.040	0.110***
Born in a NESB country	0.084	0.166***	0.053***	0.067
Metropolitan residence	0.723	0.678	0.622***	0.562***
Family characteristics at age 14				
Lived with both parents at 14	0.751	0.695	0.648***	0.525***
Mother employed at 14	0.720	0.579***	0.671	0.565***
Family income-support history				
Heavy family income-support receipt	0.275	0.387*	0.377**	0.559***
Moderate (early) family income-support receipt	0.240	0.226	0.273**	0.228
Moderate (late) family income-support receipt	0.137	0.135	0.124	0.096*
Parental characteristics				
Age of mother	47.282	47.238	46.583	44.607***
Number of children of mother	2.956	2.990	2.987	3.368***
Mother is a smoker	0.162	0.148**	0.253***	0.423***
At least one parent born in a NESB country	0.265	0.404***	0.207***	0.188***
Mother's highest education: Year 12 or equiv.	0.085	0.086	0.088	0.067
Mother's highest education: above Year 12	0.685	0.660	0.597***	0.552**
Father's highest education: Year 12 or equiv.	0.156	0.119	0.159	0.116
Father's highest education: above Year 12	0.522	0.461	0.358***	0.260***
Number of observations	1,544	452	1,200	356

Notes: Parental migrant status is reported by the young person. All other parental characteristics are reported by the parent. Unweighted. *, ** and *** denote sample means that are significantly different from the reference category at the 10%, 5% and 1% level respectively.

Table A2: Sample sizes

Number of respondents	Total	With data on marijuana use
Total in wave 2	3,623	2,480
After sample restrictions	3,560	2,439
Completed Year 12	2,748	2,023
With data on obtaining an ATAR score	2,531	1,883
Reported an ATAR score	1,826	1,440
With matched parental records	1,819	1,291
Completed Year 12	1,455	1,093
With data on obtaining an ATAR score	1,351	1,025
Reported an ATAR score	1,012	805

Note: The sample restrictions are as follows: exclude cases where the responding parent is not the young person's biological mother, where the parent has not lived with the young person in the past five years, or where a young person was still attending school.

Table A3: Sample means by middle-years experiences

	Tried marijuana by age 14		Suspension/ Expulsion	
	No	Yes	No	Yes
Outcomes				
Completed Year 12	0.870	0.605***	0.854	0.539***
Obtained an ATAR score	0.855	0.790*	0.833	0.666***
ATAR score (scale: 0-100)	76.749	70.659**	75.760	68.002***
Studying	0.648	0.475***	0.625	0.406***
Working	0.290	0.374**	0.309	0.417***
Neither studying nor working	0.062	0.150***	0.067	0.177***
Middle-years experiences				
Tried marijuana by age 14	0.000	1.000	0.047	0.251***
Suspended/expelled from school	0.123	0.491***	0.000	1.000
Other experiences				
Repeated a year	0.079	0.196***	0.087	0.177***
Number of schools	2.910	3.718***	2.892	3.576***
Own characteristics				
Female	0.542	0.511	0.532	0.316***
Indigenous Australian	0.020	0.054***	0.026	0.059***
Born in a NESB country	0.084	0.007***	0.084	0.046***
Metropolitan residence	0.690	0.649	0.676	0.627**
Family characteristics at age 14				
Lived with both parents at 14	0.785	0.549***	0.776	0.585***
Mother employed at 14	0.706	0.718	0.710	0.670**
Family income-support history				
Heavy family income-support receipt	0.243	0.392***	0.235	0.438***
Moderate (early) family income-support receipt	0.230	0.216	0.236	0.223
Moderate (late) family income-support receipt	0.081	0.084	0.084	0.070
Parental characteristics				
Age of mother	47.486	47.814	47.164	46.417**
Number of children of mother	2.901	3.037	2.938	3.014
Mother is a smoker	0.152	0.303***	0.160	0.319***
At least one parent born in a NESB country	0.264	0.179**	0.252	0.210**
Mother's highest education: Year 12 or equiv.	0.097	0.082	0.091	0.061*
Mother's highest education: above Year 12	0.649	0.730*	0.648	0.678
Father's highest education: Year 12 or equiv.	0.148	0.134	0.156	0.116*
Father's highest education: above Year 12	0.497	0.437	0.492	0.315***
Number of observations	2,235	204	2,829	725

Notes: Parental migrant status is reported by the young person. All other parental characteristics are reported by the parent. Unweighted. *, ** and *** denote sample means that are significantly different from the reference category at the 10%, 5% and 1% level respectively.

A.2 Additional details on methodology

A.2.1 The basic regression model

The results presented in Section 5 are obtained by estimating reduced-form regression models of the following type:

$$Y_i = \alpha_i + \beta_E E_i + \beta_X X_i + \varepsilon_i \quad (1)$$

where *i* indexes individuals, E_i is a vector of middle-years experiences for individual *i*, and X_i is a vector of controls for own, family and parental characteristics of individual *i*. α_i, β_E and β_X are parameters to be estimated, with β_E capturing the total direct effects of the middle-years experiences on the outcome in question, holding constant other observable factors.

Equation 1 is estimated as a probit model for binary outcomes (Year 12 completion and obtaining an ATAR score), a multinomial logit model for multi-categorical outcomes (economic activity) and an ordinary least squares (OLS) model for linear outcomes (the ATAR score).

A methodological issue arises in estimating equation 1 for two of the outcomes because data on obtaining an ATAR score and the ATAR score are not available for all observations. This is because students can only obtain an ATAR score after completing Year 12, and the ATAR score is only available when a student obtains one. Further, those who complete Year 12 (or obtain an ATAR score) are unlikely to be a random sub-sample of the overall YIF sample.

A standard way to address this issue is to estimate these two outcomes explicitly taking account of the first-stage selection process using a Heckman-style sample selection model. For obtaining an ATAR score, for example, such a model would consist of two regression equations: a selection equation where the dependent variable is Year 12 completion, estimated on all observations, and the main equation where the dependent variable is obtaining an ATAR score, estimated on those who completed Year 12. This correction technique, however, requires at least one exclusion restriction, which is a variable that influences Year 12 completion but not the decision to obtain an ATAR score, given Year 12 completion. In the case where the outcome is the ATAR score, the exclusion restriction is a variable that influences Year 12 completion and obtaining an ATAR score, but not the ATAR score itself, given a score is obtained. Unfortunately, in this context it is not possible to find a variable that plausibly influences one educational outcome without affecting another closely related educational outcome. In other words, we cannot find a valid exclusion restriction for either outcome.

Therefore, the regressions for obtaining an ATAR score and the ATAR score itself are estimated on selective samples (those who have completed Year 12, and those who obtain an ATAR score, respectively). While the results from this estimation approach hold for the respective samples, they cannot be straightforwardly generalised to the total population. But it is possible to speculate about what the population parameters might be in this case. Since the factors that influence the probability of obtaining an ATAR score are likely to be correlated with those that influence the probability of completing Year 12, the effects of the explanatory variables on the probability of obtaining an ATAR score for the total population are likely to be higher (in absolute value) than the effects for the selective sample. Likewise for the effects on the ATAR score. In other words, the estimates reported in Tables 4 and 5 in the main text are likely to be conservative estimates of what we might were we able to estimate these models for the whole cohort.

A.2.2 The potential endogeneity of middle-years experiences

Another methodological issue is that the middle-years experiences are potentially endogenous. Thus, while a significant negative relationship between a middle-years experience and a later outcome can show that those who have the experience are likely to have a poorer outcome, it is unable to determine whether or not the experience per se leads to the poorer outcome. Endogeneity may arise from omitted variables (or selection on unobservables), where both the middle-years experiences and later outcomes are driven by the same unobserved factors. Such unobserved factors could be individual idiosyncratic characteristics (such as motivation) or family idiosyncratic characteristics (such as parental support). This is related to the issue of selection bias discussed above, although in this case it is those who have the experience that may be a non-random subset of the population. Ignoring the potential endogeneity of the middle-years experiences may lead to biased estimates of their impact on later outcomes.[44]

One way to address this kind of endogeneity is to use an instrumental variable (IV). A valid IV is one that is correlated with the middle-years experience but not correlated with the unobserved factors that influence the outcome. Instruments for drug use that have been used in the literature include cocaine prices, marijuana decriminalisation, penalties for marijuana possession, and drug possession arrest rates or smoking status. However, the validity of these instruments is sometimes questionable (see the critique in French and Popovici (2011)).

[44] Another potential identification problem – reverse causality – can be ruled out in this case, as the experience (e.g. trying marijuana by age 14) precedes the outcome (e.g. completing Year 12).

Moreover, there is unlikely to be a variable that is correlated with suspension/expulsion (an educational experience) without being correlated with unobserved determinants of an educational outcome (e.g. Year 12 completion). Therefore, an IV approach is not considered in this study.

In the absence of an IV, a method developed by Altonji et al. (2005) can be used to assess the sensitivity of the estimates to varying degrees of correlation (i.e. degrees of endogeneity) between the unobserved determinants of the experience and the outcome in question. This study follows such an Altonji et al.' approach in two steps. The first step is based on a simultaneous equations model containing two equations, one for the middle-years experience and the other for the outcome. The model is estimated without any exclusion restrictions but with the correlation coefficient of the error terms of the two equations being constrained to a certain value. Various values of the correlation are considered, which allows one to examine how sensitive the effect of the experience on the outcome is with respect to such changes. This analysis shows the threshold of selection on unobserved factors at which the experience no longer has a statistically significant effect on the outcome. The second step involves calculating the ratio of selection on unobservables to selection on observables that would be required to completely explain the observed association between middle-years experiences and later educational outcomes. The advantage of this Altonji et al.'s method is that it allows one to informally gauge the extent of a causal relationship, if any, between a potentially endogenous variable and the dependent variable, without requiring an exclusion restriction.[45]

Table 7 in the main text presents results from constrained simultaneous equations models, which take the form of bivariate probit models for binary outcomes (Year 12 completion, obtaining an ATAR score, studying at age 20) and systems combining a linear regression and a probit regression for linear outcomes (the ATAR score). The constraint is that the correlation coefficient of the error terms (ρ) in each system of regressions is imposed to range from -0.1 to -0.5. Negative values for the correlation are used based on the assumption that unobserved factors that have a positive effect on marijuana usage and suspension/expulsion

[45] There are other methods to address endogeneity. One method uses a fixed-effects model to account for unobserved individual characteristics. The fixed-effects model assumes that individual heterogeneity is constant over time and thus can be differenced out using repeated observations of the same individual. While the YIF survey has two waves, wave 1 began when the middle-years experiences had long taken place (those experiences are thus constant across waves). Therefore, the YIF data cannot be used to examine the impact middle-years experiences on later outcomes based on a fixed-effects model. The second method uses sibling data to control for unobserved family characteristics. This method assumes that family characteristics are the same for siblings. Therefore, differences in sibling outcomes must be due to differences in explanatory variables other than family characteristics. However, sibling data are not available from the YIF survey.

are likely to have a negative effect on later outcomes, i.e. assuming selection bias is negative. Assuming a zero ρ is equivalent to estimating a single regression of each outcome treating the middle-years experiences as exogenous. Higher values (in absolute terms) of ρ indicate higher correlation in the unobserved factors that shape both the outcome and the middle-years experience.

Complications arise since for each outcome, there are two potentially endogenous explanatory variables: marijuana usage and suspension/expulsion from school. It is difficult to evaluate the extent of selection bias by simultaneously considering more than one endogenous variable without making assumptions about the correlation between unobserved determinants of the endogenous variables and of the outcome and each endogenous variable. Thus, to evaluate the extent of selection bias due to each potentially endogenous variable, we study its impact on the outcome when the other endogenous variable is excluded from the regression. This relegates the other middle-years experience to the group of unobservables. Given that the Altonji et al.'s (2005) method is designed to tease out causal relationships from the influence of unobservables, this does not lead to any additional difficulties in interpretation. Each simultaneous equations model controls for all other variables included in the 'full' specifications in Tables 3-5.

Comparing Table 7 with Tables 3-5, it is clear that for all educational outcomes, the effect of the experience in question is stronger when the other experience is not controlled for. For example, when $\rho = 0$, which is equivalent to estimating a single-equation regression, the marginal effect of marijuana usage on the probability of Year 12 completion is -20 percentage points when suspension/expulsion is excluded (Table 7). This estimate is twice the estimate when suspension/expulsion is included in the regression (Table 3, column 3), which suggests that the included experience at least partly absorbs the effect of the excluded experience.

Table 8 in the main text shows results from the second step in the Altonji et al. (2005) process, i.e. calculating the ratio of selection on unobservables to selection on observables that would be required to completely explain the observed association between middle-years experiences and later educational outcomes. Altonji et al. suggest that a ratio of estimate to bias of between 0 and 1 is indicative of an association between two variables that can plausibly be entirely explained by selection bias, whereas a ratio of estimate to bias of greater than 1 is likely to indicate an association at least partly driven by a causal relationship. The results presented in Table 8 show that in most cases the ratio of estimate to bias falls between

0 and 1. In other words, most of the significant effects found in Tables 3-6 can plausibly be explained by selection bias, even for the association between suspension/exclusion and Year 12 completion which until now has appeared the most likely to be at least partly driven by a causal relationship.

There are two exceptions to this conclusion, however, where the ratio of estimate to bias is *negative*, which reflects the ratio of negative coefficients to positive implied biases. The implication is that young people who report using marijuana in the middle years have other characteristics that actually *increase* the probability of obtaining an ATAR score and the ATAR score obtained. If this interpretation is correct then the estimated coefficients on marijuana use in Tables 4 and 5, and in the first row of Table 8, will actually *underestimate* the causal effect of marijuana use on obtaining an ATAR score and the ATAR score obtained. In other words, middle-years marijuana use *reduces* the probability of obtaining an ATAR score by at least 8.2 percentage points (using the estimate in Table 4) or 10.1 percentage points (using the estimate in Table 8). Similarly, middle-years marijuana use *reduces* the ATAR score obtained by at least 3 points (using the estimate in Table 5) or 4.7 points (using the estimate in Table 8). In the case of obtaining an ATAR score, the underestimated coefficient is statistically significant in the Table 8 model (excluding suspension/expulsion) but just below statistical significance in the Table 4 model (including suspension/expulsion). It seems likely, therefore, that the 'true' causal relationship, were we able to estimate it, would be statistically significant at least at the 10% level. For the ATAR score obtained, neither underestimated coefficient is statistically significant, although again were we able to estimate the 'true' causal relationship it seems plausible that it would be statistically significant at least at the 10% level.

Is this interpretation correct? Van Ours and Williams (2009) suggest something similar using different Australian data and looking at the relationship between early-adolescent marijuana use and years of schooling. And in the YIF data we can see some observable characteristics of marijuana users that are positively associated with educational outcomes (e.g. mother's education level), which may indicate similarly positive unobserved characteristics (see Table A3). This interpretation is also consistent with the increase in the estimated association between marijuana use and the probability of obtaining an ATAR score when we condition on observables in Table 4, although there is no similar increase in association with marijuana use in the model for ATAR score obtained (where suspension/expulsion is also controlled for).

Ultimately, however, the confidence with which we draw this particular conclusion – that middle-years marijuana use has a negative causal impact on obtaining an ATAR score and on ATAR score obtained – must be tempered by concerns over small sample size in the YIF data once we discard unmatched questionnaires and once we condition on Year 12 completion (and for ATAR score obtained, on obtaining an ATAR score). There are very few middle-years marijuana users who complete Year 12 and who obtain an ATAR score in the YIF sample we use to estimate the relevant models, so we should be cautious in generalising from these particular estimates.

A.2.3 Differential impacts by SES

Finally, in the Section 7 analysis of whether associations between middle-years experiences and later outcomes vary by SES, particular care must be taken in generating the correct marginal effects in the models where interaction terms are included in Equation 1 (see Norton et al., 2004). Here, marginal effects are calculated for the interaction terms as described in Ai and Norton (2003) and Karaca-Mandic et al. (2012). Note, however, that these marginal effects are only reported for three of the four outcomes (Year 12 completion, obtaining an ATAR score, and ATAR score).