NATSEM

National Centre for Social and Economic Modelling
• University of Canberra •

APPSIM – Modelling Education

Alicia Payne, Richard Percival and Ann Harding

Working Paper No. 11
June 2008

About NATSEM

The National Centre for Social and Economic Modelling was established on 1 January 1993, and supports its activities through research grants, commissioned research and longer term contracts for model maintenance and development with the federal departments of Family and Community Services, Employment and Workplace Relations, Treasury, and Education, Science and Training.

NATSEM aims to be a key contributor to social and economic policy debate and analysis by developing models of the highest quality, undertaking independent and impartial research, and supplying valued consultancy services.

Policy changes often have to be made without sufficient information about either the current environment or the consequences of change. NATSEM specialises in analysing data and producing models so that decision makers have the best possible quantitative information on which to base their decisions.

NATSEM has an international reputation as a centre of excellence for analysing microdata and constructing microsimulation models. Such data and models commence with the records of real (but unidentifiable) Australians. Analysis typically begins by looking at either the characteristics or the impact of a policy change on an individual household, building up to the bigger picture by looking at many individual cases through the use of large datasets.

It must be emphasised that NATSEM does not have views on policy. All opinions are the authors' own and are not necessarily shared by NATSEM.

Director: Ann Harding

ISSN 1834-7630
ISBN 9781740882934

National Centre for Social and Economic Modelling
University of Canberra ACT 2601 Australia
170 Haydon Drive Bruce ACT 2617

Phone + 61 2 6201 2780 Fax + 61 2 6201 2751
Email natsem@natsem.canberra.edu.au
Website www.natsem.canberra.edu.au

Abstract

With the partnership of the Australian Research Council and 13 Commonwealth departments, NATSEM is developing a dynamic microsimulation model to project Australia's future – the Australian Population and Policy Simulation Model (APPSIM). The model will 'age' a simulated Australian population by forecasting the life paths of individuals and families from 2001 to 2050, providing a tool for the analysis of the distributional and revenue impacts of policy into the future. A key module of the APPSIM model is education, which will model individuals' educational participation and qualifications in school, TAFE and university studies.

This working paper provides a review of the methods used to model education in existing dynamic microsimulation models from around the world, and an initial proposal for how Australia's education system could be modelled in APPSIM. Available data on education in Australia is also reviewed.

Author note

Alicia Payne is a Senior Research Officer at NATSEM. Professor Ann Harding is the Director of NATSEM and Associate Professor Richard Percival is a Principal Research Fellow.

Acknowledgments

The authors would like to gratefully acknowledge the funding provided by the Australian Research Council (under grant LP0562493) for the development of APPSIM, and the 13 research partners to the grant: Treasury; Broadband, Communications and the Digital Economy; Education, Employment and Workplace Relations; Health and Ageing; Finance and Deregulation; Families, Housing, Community Services and Indigenous Affairs; Innovation, Industry, Science and Research; Immigration and Citizenship; Prime Minister and Cabinet; the Productivity Commission; Centrelink; and the Australian Bureau of Statistics.

The authors would especially like to thank Cezary Kapuscinski, and the Department of Education, Science and Training, for meeting to discuss the data requirements of this module.

The authors would also like to especially thank those who have developed the models reviewed in this paper for their generosity with time and resources to give us insight into how their models work and what they have learnt: Rick Morrison, Chantal Allen, Eden Cloutier and Keith Wilde from the Department of Human Resources and Social Development, Canada; Geoff Rowe from Statisitics Canada; Dennis Fredriksen from Statistics Norway and Melissa Favreault from the Urban Institute.

The authors would also like to thank Clair Matthews for editorial comments.

General caveat

NATSEM research findings are generally based on estimated characteristics of the population. Such estimates are usually derived from the application of microsimulation modelling techniques to microdata based on sample surveys.

These estimates may be different from the actual characteristics of the population because of sampling and non-sampling errors in the microdata and because of the assumptions underlying the modelling techniques.

The microdata do not contain any information that enables identification of the individuals or families to which they refer.

Contents

1 Introduction

Dynamic microsimulation models provide an effective tool for governments to project the size and characteristics of the population decades into the future and the expenditure and revenue associated with current and proposed policy in this projected environment (see Harding and Gupta 2007 and Gupta and Harding 2007). Dynamic microsimulation models are currently being used by governments for this purpose in many countries, including the UK, US, Canada, Sweden and Norway. The current concern of government about the ageing of Australia's population over coming decades makes the need for such a tool to estimate future population and policy particularly apparent. In the 1990's NATSEM developed an Australian dynamic microsimulation model called DYNAMOD. Australian Research Council funding, with the partnership of 13 Commonwealth departments and the availability of longitudinal data in the Household Income and Labour Dynamics of Australia (HILDA) Survey have given NATSEM the opportunity to develop a new dynamic model to project Australia's future – the Australian Population and Policy Simulation Model (APPSIM).

Beginning in 2001 and projecting out to 2050, APPSIM will model the life 'events' of Australians over this period. The base population of the model is based on the 2001 Census and, as the model 'ages' the population, individuals will be born and added to the population, while others will die and leave the simulation. With each passing year each individual passes through a series of modules that predict events that will happen to them in that year. APPSIM's modules will include demographics (births, deaths, immigration and emigration), household formation and movement (marriage, cohabitation, children leaving home, divorce and separation), education and training, labour force, earnings, housing, other income and expenditure, household assets and debt, social security, taxation and health and aged care. Thus the model attempts to capture all life events and characteristics of Australians that pertain to policy planning. The information about individuals modelled in each of the modules can be used by other modules and can be output for its own sake. For example, education and training characteristics impact on a range of factors being modelled in other modules, including fertility, labour force participation, and earnings.

The purposes of this paper are to discuss issues and research relating to the development of APPSIM's education module and to propose how the module will work in a simple form. Education is a vital aspect of the model because of the broad policy impacts of education. An individual's educational participation (both 'current' and historical) has implications for labour force participation, earnings, fertility, social security and taxation.

Section 2 of this paper summarises a review of how existing dynamic microsimulation models have approached the modelling of education in order to inform the development of APPSIM's education module. Section 3 reviews available data on education in Australia. Section 4 outlines the proposed approach to modelling education in APPSIM.

2 Review of education modules of dynamic microsimulation models

2.1 DYNAMOD

DYNAMOD is a dynamic microsimulation model previously developed by NATSEM. DYNAMOD-2's education module operates on an annual basis. In January each year transitions into, out of and through the levels of the education system are modelled on the basis of transition probabilities.

DYNAMOD assumes full-time education to be compulsory for all people aged 6 to 15 years, and a child begins school when they reach the age requirement of their state. Transition probabilities are used to allocate children starting school to a Government, Catholic or independent school. Students are then assumed to remain in school till year 10.

Transition probabilities determine whether students then continue on to year 11, and for students that continue, whether they will complete year 12 also. After year 10, TAFE also becomes an option for anyone over 15. The probabilities are derived from *The Australian Youth Survey*. It is assumed that students do not repeat years or take years off.

Tertiary courses at TAFE or university are assumed to take a certain number of years, which differs depending on whether the student is studying full or part-time. For those that commence a tertiary course the decision to drop out or continue is modelled each year until the number of years to complete the course has been reached. The decision to study full or part-time is also modelled annually.

Entrance into university is open to people who have completed year 12 or a TAFE course, and those aged over 21 who can enter as mature age students. These transitions have been estimated as logit equations on the basis of *The Australian Youth*

Survey. Explanatory variables for this transition include age, sex, ethnicity, parents' occupation, income and level of education, type of school attended, state of residence, whether the student has a disability, the rate of participation in the level and other economic conditions in the state.

For those who complete an undergraduate university degree, further tertiary eduction is modelled with the option of a graduate diploma, a masters or a higher degree. A graduate diploma requires people to have done a bachelor degree, while a masters requires either a bachelor degree or graduate diploma. Having completed a graduate diploma or masters, people can be allocated to do a higher degree.

Commencement of tertiary courses is based on published administrative data from the then Department of Employment Education and Training (*Selected Education Statistics)* and the National Centre for Vocational Education Research (*Selected Vocational Education and Training Statistics).* The main variables used are age, sex, highest educational qualification and study type (full or part-time). Attempts were made to reproduce the aggregate stocks by age in any given year of study in a course (Robinson and Bækgaard 2002, p3).

It is assumed that people will not do a 'lower' degree than the degree that they have already completed. For example someone with a masters would not do a second bachelor degree. It is also assumed that those with university training will not enrol in a TAFE course.

Figure 1 and Table 1 show the sequence of decisions, and who is 'at risk' of certain events.

Modelling labour force participation and earnings while studying

In DYNAMOD-2 the working hours and earnings of full-time students are also modelled as part of the education module. The earnings of people who are employed full or part-time and are *not* full-time students are modelled on the basis of data from the ABS *Income Distribution Survey*. Student's working hours and earnings are modelled separately as they are more likely to vary across the year.

The model estimates three things: first the 'employment decision' of whether to work or not, the student's average hourly wage rate, and last the annual number of hours worked. Students are first divided into those employed and those not employed in the initial year, before their employment for the next year is modelled.

Figure 1 **Transitions and Explanatory Variables in DYNAMOD's Education Module**

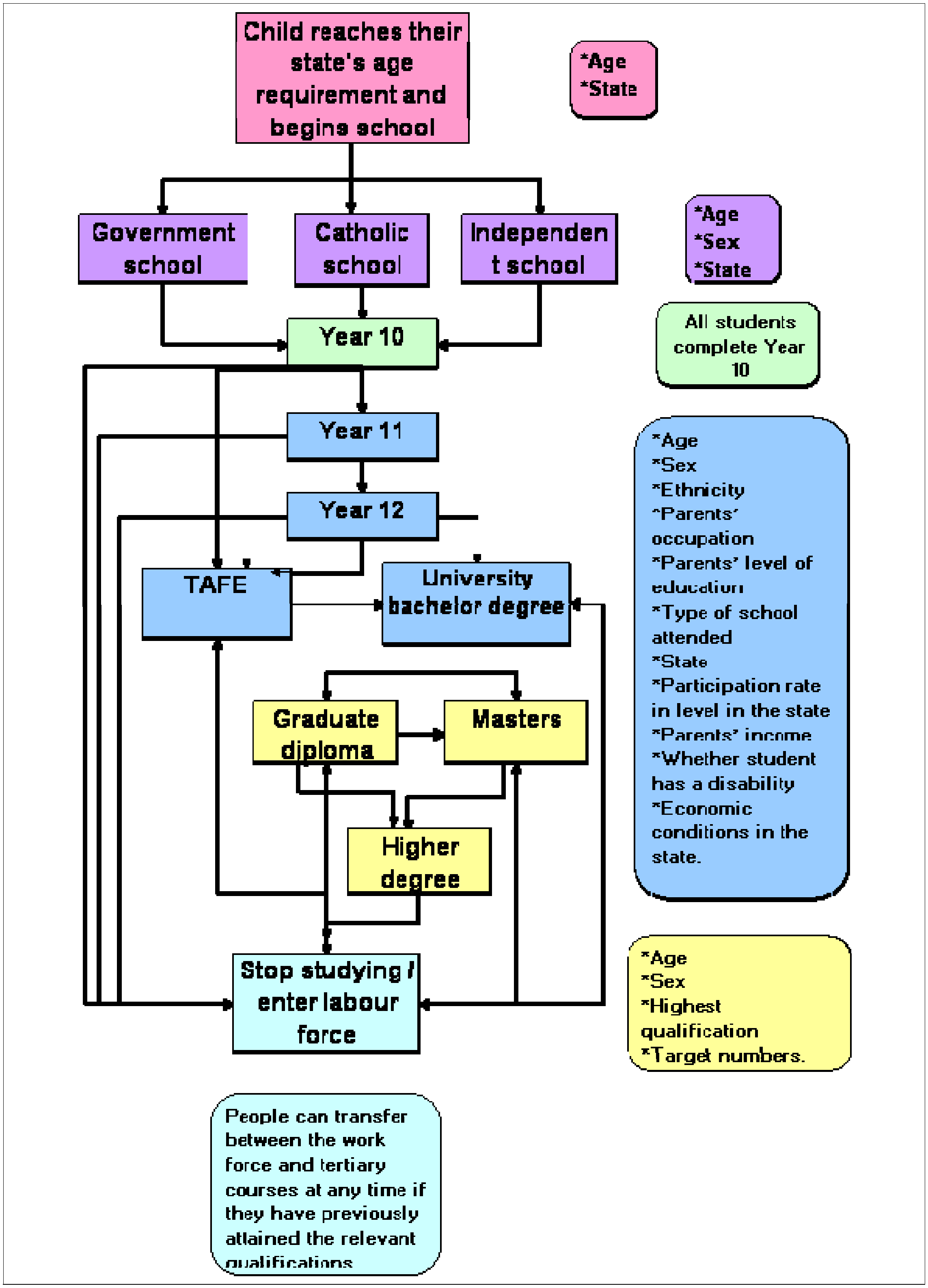


Table 1 **Modelling Education Decisions in DYNAMOD**

Educational Decisions (In sequence)	**Who is 'at risk' of commencing**	**Explanatory Variables**
Commencing Primary school, will continue to Year 10	All children in the base data, those 'born' in the model and those entering through immigration who are between the ages of their state's school age and 15 years (when they complete year 10)	Age and state of residence (for minimum age)
Government, Catholic or Independent School?	All children beginning or continuing school. Although all who begin school continue to Year 10, movement between sectors is modelled.	Age, sex, state and the year levels that they have completed and will enter
Year 11, Year 12, TAFE, First Year University	It is assumed that no breaks are taken between Year 10, 11 and 12. Thus all those who have just completed the previous school year (in the previous calendar year) are eligible to begin the next year. TAFE can be commenced by anyone who has completed Year 10 or higher and first year university by anyone who has completed Year 12, a TAFE course or is over 22 years of age.	Age, sex, ethnicity, parents' occupation, parents' level of education, type of school attended, state of residence, participation rate in level in the state, parents' income, whether student has a disability and economic conditions in the state.
Continuing years in higher education: full or part-time	The decision to continue for another year or drop out is modelled for all who have commenced a course at uni or TAFE until they have reached the number of years required to complete the course.	The target number of males / females undertaking the course.
Post graduate study: graduate diploma, masters or higher degree: full or part-time	Anyone who has completed a bachelor degree can commence a graduate diploma or masters, and anyone completing one of these can do a higher degree	Age, sex, highest qualification and the target number.

Source: Hardin, 1994.

Modelling the employment choice

Whether or not a student works is modelled on the basis of observable characteristics such as age, sex, marital status and type of course the person is studying, and the unobservable factor of the level of utility gained from employment conditional on job availability (Robinson and Bækgaard 2002, p6). Based on this whether or not a student works is modelled using an *index function model* to estimate their net utility from employment y *, such that:

$$y^* = \beta'X + \varepsilon$$

Where $\beta'X$ is the index function (X is a vector of observed explanatory variables for the model and β is a vector of coefficients) and ε is a random disturbance term. Since y^* is unobserved and we observe only whether a person is employed (when $y = 1$) or not (when $y = 0$), then:

$y = 1$ if $y^* > 0$, and

$y = 0$ otherwise.

Then, the probability that $y = 1$ and the person is employed is expressed

as:

$P(y = 1) = P((\beta'X + \varepsilon) > 0)$

$= P(\varepsilon > - \beta'X)$

$= 1 - F(-\beta'X)$

where F represents the logistic function which is symmetric. Hence we

can write:

$P(y = 1) = F(\beta'X)$

where:

1 exp(′X)= exp(β′X) /1+ exp(β′X)

(Robinson and Bækgaard 2002, p6).

The employment decision is formulated as a two-period transition model based on the employment situation in two consecutive years. It is possible to identify four key transitions, namely, flows into and out of employment and non-employment.

While we are interested in all four of these transitions, we need estimate only the probability of one transition from each group in year (t - 1), since the probability of the other transitions will be the residual. For each of the two groups, our interest lies in estimating the probabilities of the transitions into work in the current year:

1. P(student employment (t) | student employment (t-1))
2. P(student employment (t) | not in student employment (t-1))

A person is deemed to be in 'student employment' if they are in full-time study for the whole financial year and work in paid employment for one hour or more in the period. (See Robinson and Bækgaard 2002, p6).

Estimating earnings for full-time students in employment

Next the hourly wage and annual number of work hours is modelled for those students that have been deemed to be working in the current year. DYNAMOD-2's documentation states that while it would normally be desirable to model the two with a joint model, tests have shown that the interdependence between the two processes is particularly simple for full-time students, and thus it is appropriate to estimate the two equations separately in two steps (Robinson and Bækgaard 2002, p7).

The hourly wage rate is estimated on the basis of a number of personal and demographic variables that have been shown to be highly correlated with the wage rate (Robinson and Bækgaard 2002, p7). Students are again divided into two groups:

- Full-time students who were employed in the previous year; and
- Those who were not full-time students in the previous year, or those that were full-time students but were not employed

This is done to include the previous year's wage rate in the equation for those that were full-time students who were employed in the previous year. This division remains for the modelling of annual number of hours. Regression models are used to estimate annual hours for each group separately (see Robinson and Bækgaard 2002, p9 for details of the equations).

2.2 SAGE

The SAGE model ('Simulating Social Policy in an Ageing Society') was developed by the London School of Economics and King's College London, UK. The aim of the model is to be a tool 'with which to make evidence-based social policy projections of the circumstances of older people for the twenty-first century' (Scott and Zaidi, 2004:1). The model's focus on older people has implications for the modelling of education, given that the key point of interest is modelling the pension entitlements of older citizens. Educational qualifications influence labour force participation and thus life -long income and entitlement to the aged pension in the UK, but the details of the timing of entries into and exits from education are not a key focus of SAGE.

Education is modelled in SAGE as a 'discrete phase of the life course' (Scott and Zaidi 2004, p1). Once a person has exited education into the workforce, they do not re-enter education, which is assumed to be completed by the age of 22 years (Scott and Zaidi 2004, p6). 'Gap-years', returning to further qualifications and dropping out of qualifications are not modelled. The empirical data presented in the beginning of Scott and Zaidi's 2004 technical paper shows that many people further their qualifications in their mid-twenties, but that only a small proportion of people do so after the age of 28. To account for this the rate of each level of educational qualification of people up to 28 years of age is imputed onto those aged under 22 (Scott and Zaidi 2004, p6-7).

The rates are also adjusted for differences in cohorts. One such example of a cohort effect is that in the UK the minimum-age to leave school was raised from 14 to 15 years in 1947, and then to 16 in 1972, raising the level of education of people as they were affected by these changes. There is also a trend towards higher qualifications in younger cohorts. For APPSIM it will be important to take account of any such changes in the Australian system.

SAGE's education module attributes only four levels of education:

1. No qualification;
2. General Certificate of Secondary Education (GCSE, equivalent to year 10 in Australia);
3. General Certificate of Education A level (equivalent to year 12 in Australia); and
4. Higher qualifications.

The model does not differentiate between graduate or higher degrees, or between university or other post-school qualifications. Qualifications are attributed based on the person's age at leaving education (under 22). The education decision is made each year at the person's birthday between their 16th and 22nd birthdays. Based on age, sex and the 'social class of parent' (Scott and Zaidi 2004, p3) the decision is made sequentially between 'unqualified student', 'unqualified adult', 'qualified student' or 'qualified adult', as follows:

From 0 to 15 all records are unqualified students

↓

At 16 and then 17 (the minimum-age to leave school) a proportion are attributed to unqualified adult (those leaving without GCSE)

↓

At 18 some of those remaining in education leave and are attributed to qualified adult (with their A levels)

↓

At 19 and 20 some are attributed to qualified student, as they continue their education.

↓

At 21 remaining students become qualified adults. At this point there are no more unqualified or qualified students in the simulation – only qualified or unqualified adults (Scott and Zaidi 2004, p8).

2.3 DYNACAN

DYNACAN is a dynamic microsimulation model developed by the Department of Human Resources and Social Development, Canada. In DYNACAN, years in education and level of qualification are modelled simultaneously, based on the technique used in CORSIM (Morrison and Dussault 2000, p16). It is assumed that once someone leaves education they do not return. In the adaptation of the CORSIM method for DYNACAN, the need to adjust for people returning to further their qualifications became apparent. To address this, qualifications are 'instantly' adjusted at the age of 32. Five levels of qualifications are attributed: high school, trade or vocational studies, college, undergraduate university, and graduate studies. The variables used to model the level attained include sex, age, parents' education, living arrangements, whether parents' own their home, marital status, and whether or not the person has children.

2.4 MOSART

MOSART has been developed by Statistics Norway and used extensively in reforming the Norwegian public pension system. MOSART does 'recursive' simulation, where events are simulated one at a time in a given order (in one period) with conditional transitional probabilities (Fredriksen 1998, p13). It is intended that with further development MOSART's education module will be used for projecting labour supply by educational characteristics, but it currently projects education for use as an explanatory variable[1]. The module attributes two educational characteristics to people: their current educational activity and their highest attainment.

Educational characteristics are simulated annually for each person in the following sequence of decisions:

> Will the person's educational activity of the previous year (if any) be completed this year in the spring semester?
>
> ↓
>
> Will the person start or continue as a student in the autumn semester this year?
>
> ↓
>
> If the person is allocated to study, what will be their field of study?
>
> ↓
>
> Given the subject, at what 'level' will they be studying?

There is no age limit to when a person can return to begin new study. Fredriksen (1998) explains that dividing the module into these four steps, each with a limited number of possible outcomes, makes estimation and policy analysis easier than less steps with more possible outcomes. It is also easier to update, as some steps may not need to be changed.

MOSART's education module attributes educational attainment using transition probabilities based on gender, age and educational participation and attainment in the previous year (Fredriksen 1998, p 57). Men and women are modelled separately. Transition probabilities also take account of aggregate factors impacting on educational trends, such as 'unemployment, relative wages and capacity in the educational system' (Fredriksen 1998, p. 57).

Transition probabilities are also adjusted to meet yearly targets of the number of students by gender and age (Fredriksen 1998, p16). Fredriksen states that a proportional adjustment for this purpose would not suffice as many of the probabilities are close to one, especially among those who were students in the previous year (1998:58). Instead they are adjusted by the following, where p_i^0 is the base line probability of becoming or continuing as a student, and r is the adjustment factor:

---

[1] Indicated in an email from Dennis Fredriksen at Statistics Norway who works on the model, 12 January 2007.

$$p_i^{adjusted} = p_i^{0} \cdot (1+r)/(1+p_i^{0} \cdot r),$$
$$r \geq -1$$

This can be interpreted as changing the constant term of a logit function.

> 'The adjustment factor 'r' is calculated 'by iteration over first order Taylor polynomial each year. ... This is based on (the assumption that) the probabilities $p_i^{adjusted}$ are a function of 'r' and that the number of events equals the sum of the transition probabilities with the drawing method[2] in the MOSART model. For small probabilities the adjustment by (the equation above) is quite similar to proportional adjustment.' (Fredriksen 1998, p. 58).

For example, a different 'r' value is used for men and women, for particular years. Fredriksen notes that a 'weakness' of the method used in MOSART is the focus on the highest attainment level. He explains that in Norway, it is common for people to 'repeatedly' change their field of study and to complete their studies in segments punctuated by years off. A focus on the characteristic of highest attainment, rather than educational history, or earlier education, when using education to model other events (for example, field of employment) can mean that the number of people in certain professions is over or under-estimated. The nature of variables output in the education module is a key point to consider in the design of APPSIM as they are central to several other modules.

2.5 SESIM

Sweden's Ministry of Finance has developed a dynamic microsimulation model called SESIM. In SESIM events are simulated in a sequence and updated each year. Education is the second module in the model's sequence, following the modelling of demographic events. The education module of SESIM attributes four levels of education:

- Primary school (not more than 2 years of high-school education);
- High school degree (more than 2 years of high-school);
- Bachelor's degree (not more than three years of education after high-school); and
- University degree (more than three years of education after high-school).

(Pylkkänen 2000, p8).

[2] 'Mean constrained drawing methods', see Fredriksen 2003, p. 16.

All people have 'child status' up until the age of 15 and are assumed to attend school for all the relevant years up to this age (Flood et al. 2003, p5). From 16 onwards people are 'at risk' of being attributed to different educational situations and, finally, levels of attainment (Flood et al. 2003, p32). The modelling of educational events is modelled in the following way:

Event/outcome: starting high school
At risk: 16-year old not disability pensioner
Algorithm: all individuals start high school at 16 years of age.
Comment: at most, three years of studies are required in order to obtain a high school degree.

Event/outcome: exit from high school
At risk: students at high school
Model: logistic regression
Covariates: sex, parent's highest education, parent's highest age, nationality, indicator for divorced parents, household's income (quartiles), number of children in household
Comment: for those individuals who are predicted to exit, a time for this exit is also predicted by a random draw.

Event/outcome: starting university studies directly after high school
At risk: 19 years old who finished their high school degree the preceding year and who are not disability pensioners or on parental leave.
Model: logistic regression
Covariates: sex, nationality, indicator for living with parents, indicator for own children, indicator for living in big city, rate of unemployment

Event/outcome: exit from university studies before a degree
At risk: students at the university
Model: logistic regression
Covariates: sex, nationality, marital status, age, number of children, indicator for living in big city
Comment: university studies are assumed to continue three or four years. Students that exit after the third year obtain a degree, after the forth year they obtain a degree by default.

Event/outcome: transition from labor force to university studies
At risk: individuals with a degree from high school who do not have a student status
Model: logistic regression
Covariates: sex, nationality, GNP-growth, time since the high school degree, age, marital status, number of children

Event: transition from labor force to ”komvux” (high school for adults)

At risk: individuals between 20 and 64 years of age without a degree from high school, who are not disability pensioners or on parental leave
Model: logistic regression
Covariates: sex, nationality, GNP-growth, marital status, time since the high school degree, number of children, age
Comment: studies at ”komvux” are assumed to continue for three years, this gives a high school degree.

Event/outcome: university studies directly after ”komvux”
At risk: individuals who finished ”komvux” the preceding year and who are not disability pensioners or on parental leave.
Model: logistic regression
Covariates: sex, nationality, age, marital status, number of children, indicator for living in big city, GNP-growth

(Flood et al. 2003, p32-33).

2.6 DYNASIM

DYNASIM has been developed by the Urban Institute, USA. DYNASIM allocates events to each person on a yearly basis. A person's progression though education is modelled using ten 'separate cross-tabulations', based on race, sex, parental education, and the amount of time that the person has spent in school so far (Favreault and Smith 2004, p10). Race, sex and usually educational attainment groups are modelled separately.

It is possible for people to repeat grades in primary and secondary school, and to stay in college for more than four years. People can also drop out of and return to schooling at several different points.

The sequence of decisions modelled is as follows:

Entry to grade school

↓

Commencement of grade 7

↓

Commencement of grade 8

↓

Commencement of grade 9+ (accounting for retention / repeating)

↓

Failure to complete 8th grade

↓

Entry into high school (for those who have completed grade 8)

↓

Annual high school continuation (including additional retention years)

↓

Dropping out of high school

↓

Entry into college

↓

Continuation of college (up to five separate years)

↓

Dropping out of college

↓

Entering graduate school

People in their late twenties do not re-enter education. However, baseline people who were enrolled at older ages can continue.

Several transition probabilities have been altered so that allocations meet targets of the age-sex-race distribution of educational attainment based on population data[3].

2.7 LifePaths

LifePaths has been developed by Statistics Canada. In contrast to many of the models outlined in this paper, LifePaths simulates events in continuous time. This means that events can occur at anytime, rather than being attributed at an annual interval. The model is constantly updating 'waiting times' to possible outcomes, based on the likelihood of certain outcomes for each person based on their characteristics. The

[3] Much of the information on DYNASIM was sourced from an email from Melissa Favreault, a developer of the model sent 28 January 2007. Notably some of the team who worked on DYNASIM's education module are currently developing one for the US Social Security Administration's POLISIM model, details of which will not be know for some months.

event with the shortest waiting time is the next event to be allocated to the person, and as the person's status changes, the waiting times for remaining events are adjusted (see Statistics Canada 2007, p3-4). Thus education is also modelled in this way.

'Primary-secondary' and 'post-secondary' school are modelled separately, a key function of the primary-secondary module being to allocate access into the post-secondary system. Individuals begin primary school in September of the year they turn six, and continue until at least their tenth birthday. After their tenth birthday, students begin to be allocated to one of three pathways:

- To graduate without dropping out;
- To drop out, return, then graduate; or
- To drop out, return, then drop out again and not complete secondary school.

(Statistics Canada 2007, p15)

Notably, while students can drop out from the age of ten, only those of 15 or over can graduate. Each step of the process of modelling primary-secondary school involves two competing events for students: to graduate or to drop out. Thus there are three outcomes for the student: to graduate, to drop out, or to continue in school. The probabilities of these events are based on age, sex, month of birth, province of residence, calendar month and educational history. In most cases, students who drop out will return to school. When a student drops out their waiting time before re-entry depends only on their sex and province. Once they re-enter they are treated the same as the other students, although their graduation and drop out probabilities are different to those of students who have not dropped out.

This process operates within several rules:

- Students can only graduate at the end of June, August or December;
- Students who have dropped out cannot return to school in one of these graduation months;
- The chances of someone who has dropped out re-joining are highest at the beginning of term and diminish towards graduation;
- The student's year of graduation is governed by their province of residence, as the year of graduation (i.e., 11, 12 or 13) differs between the Canadian provinces; and

- Any student who continuously attends primary-secondary until age 25 (i.e., someone who has not been allocated the event of graduation or dropping out) is allocated to 'non-completion' of primary-secondary and exits the system.

The post-secondary education path that a person will take is decided when they graduate from secondary school. The possibilities include:

- Remaining a secondary school graduate;
- Various combinations of trade vocational school certificates;
- Other non-university certificates;
- University degrees; or
- Taking on, but dropping out of university or college

(Statistics Canada 2007, p19)

Overall there are 'thirty distinct fates' that can be allocated to a person. Once this 'fate' has been allocated further 'fates' are also decided. For example:

- For those who are to attend more than one institution, which one they will attend first;
- For those going to university, their province of study is allocated, influenced by their sex and their province of residence (it is assumed those studying in non-university institutions always study in their home province);
- When the province of study is known, the waiting period until the person commences the first program is calculated, dependent on province of study, the person's decided 'fate' and sex;
- One of 100 fields of study is allocated to each person attending college or university. Then a pattern of study, representing the total length of the program, and how long will be spent in full and part-time study as well as any years off is allocated.
- People attending a trade vocational institution are allocated to one of six programs. Once the type of course is chosen, the length of the program is decided immediately.

(Statistics Canada 2007, p19)

Similarly to NATSEM's DYNAMOD, LifePaths models the employment of students as a separate labour market. Full-time students are allocated to being employed, self-employed or not employed. It is first decided whether they are employed or not, and then for those 'employed' some are allocated as being self-employed. The likelihood and timing of transitions between these states is updated monthly and is influenced by sex, province, status in parental home and variables representing 'seasonal employment patterns and time trends' (Statistics Canada 2007, p32).

For the modelling of students' labour force participation, transition probabilities are estimated separately for students in primary-secondary, college, university and other schools. Similarly to DYNAMOD, labour force decisions are modelled separately for those who have and have not worked previously.

2.8 Summary of lessons for modelling education in APPSIM

Several lessons for the development of APPSIM can be drawn from this review of approaches to modelling education in dynamic microsimulation models. Table 2 compares some key aspects of the models reviewed in this paper and several lessons for the development of APPSIM's educational module can be drawn from these comparisons:

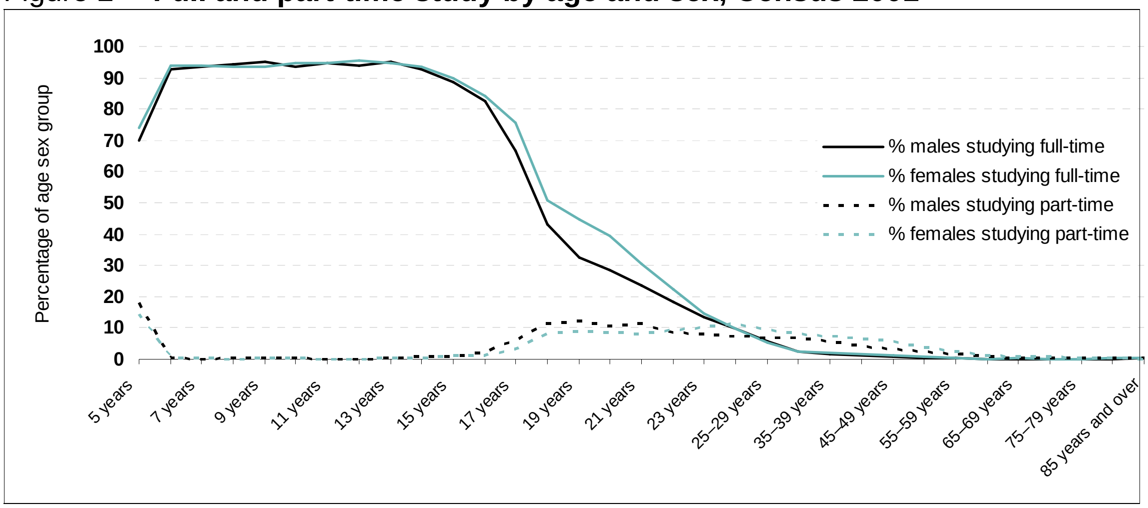
Time: It is proposed that APPSIM's education module be run at discrete (annual) time. As shown in the review of dynamic education modules in this paper, all the models except for the Canadian LifePaths model are based on annual transitions. This is also consistent with other planned modules for APPSIM.

Minimum level of compulsory education: As shown in Table 2, in the models reviewed full-time education is generally compulsory until some point during high school. In DYNAMOD, full-time education was compulsory until year 10, which is an appropriate reflection of the Australian education system. It is recommended that full-time education be compulsory up to a level of year 10 in APPSIM also. While Australian students are permitted to leave school slightly earlier than this, the significant majority stay until year 10. As shown in Figure 2, the proportion of people studying full-time only begins to decrease significantly after the age of 15 and 16 years, i.e. those likely to be in year 10 in the Australian schooling system.

Returning to education: As shown in Table 2, the models have taken varying approaches to adults returning to education. As Australians are studying more, and until later in life, combining work and study and changing career paths throughout their working lives, it would be most realistic to place the most minimal assumptions possible around people returning to education. It is proposed that the APPSIM education module will allow students to return to education at TAFE or university until a set age limit at which the sample of students in available data becomes so small that it is unable to be modelled effectively.

Data from the 2001 Census shows that while enrolments in TAFE and university peak for age groups in the late teens and early 20's, substantial numbers of students continue into older age groups (see Figure 2). The Census shows that in 2001 six per cent of men and five per cent of women aged 25-29 were studying full-time. A further seven per cent of men and nine per cent of women in this age-group were studying part-time. In older age-groups, the percentage of both men and women studying full-time drops to below 3 per cent. Greater proportions of both men and women were studying part-time into older age groups. Part-time study is likely to have less impact on incomes, labour force characteristics and fertility for example than full-time study. Thus it is proposed that the age cut-off for returning to education be set at 30, at least in the initial simple version of the module.

Figure 2 **Full and part-time study by age and sex, Census 2001**



Data source: 2001 Census Household Sample File CURF

It was also assumed in DYNAMOD that adults would only return to a 'higher' course than the highest qualification that they have completed. This assumption will also be applied in APPSIM's education module.

Table 2 **Overview of approaches to modelling education in dynamic microsimulation models**

Model	DYNAMOD	SAGE	DYNACAN	MOSART	SESIM	DYNASIM	LifePaths
Organisation	NATSEM, Australia	LSE and King's College, UK	Dept. Human Resources and Social Development, Canada	Statistics Norway	Ministry of Finance, Sweden	Urban Institute, USA	Statistics Canada
Time	Annual	Annual	Annual	Annual	Annual	Annual	Continuous
Full-time education compulsory to:	Year 10	Age 15	Age 13		Age 16	8th grade	Age 10
Restrictions on commencing education as an adult	Adults can return at any age to commence a 'higher' course	Once a person enters the workforce, which happens by the time they reach 22, they cannot recommence education. However, qualifications are 'adjusted' at the age of 28	It is assumed that when someone leaves education they cannot return. However to account for continuing education qualifications are adjusted at the age of 32.	People can return to study at any age.	People can return to school at any time. Those aged between 20 and 64 attend 'komvux' - 'university for adults'.	People in their late 20's cannot re-enter education, but 'baseline' people who were enrolled at later ages can.	
Taking years off or repeating	Up till year 12 students do not repeat or take years off. In tertiary education the decision to drop out or continue is modelled each year	Students cannot have gap years or repeat	Yes it is possible for students to drop out of education for a period and return to school / their course. Repeated years are also accounted for.	Students cannot repeat years in primary school (which goes until students are 16). They can take years off before and during a higher course, and can change into a different field of study.		People can repeat grades in primary, high school and college. It is possible for them drop out and re-enter education at several points.	From age 10 students can drop out and return to school

Model	DYNAMOD	SAGE	DYNACAN	MOSART	SESIM	DYNASIM	LifePaths
Alignment	Yes, attempts to hit targets of number by type of tertiary course by age and sex	Yes, qualifications adjusted at age 28	Yes, qualifications adjusted at 32 and numbers for each age group aligned.	Yes, adjusted to meet yearly targets of number of students by gender and age.		Yes, some transition probabilities have been altered so that output meets targets of the age-sex-race distribution of educational attainment.	
Is student employment modelled in a specific module?	Yes		No				Yes
Qualifications imputed	Year 10 Year 11 Year 12 TAFE University bachelor Graduate diploma Masters Higher degree	Year 10 equivalent (GCSE) Year 12 equivalent (A level) Higher qualification	High School Trade College Undergraduate university Graduate studies		Primary School High School Bachelor degree (no more than 3 years education after high school) University degree (more than 3 years after high school)	High School College Graduate School	Below secondary school Secondary school graduate Trade vocational school certificate Other non-uni certificates Uni degrees Uncompleted uni
Is field of study imputed?	Yes	No	Yes		Yes		Yes
Are ‘macro’ variables used in transition probabilities?	Yes, participation rates by state and economic conditions by state		No	Yes, unemployment, relative wages and capacity in the educational system			

Repeating years of education and taking years off: Assumptions about students repeating and taking years off in education has also been approached in varying ways across the models reviewed. A more detailed review of data (some yet to be obtained) will illuminate the trends current in Australian education and whether repeating and gap years are common enough to be able to be modelled, as well as whether adequate data on such flows is available. It is probable that the best approach to APPSIM will be similar to that in DYNAMOD, where years cannot be skipped or repeated in secondary school but the decision to continue or drop out is modelled each year for students in tertiary education. It is also recommended that APPSIM model the return to education for those who have dropped out of a tertiary course separately to the general workforce. This is similar to the approach taken in LifePaths. Those who have dropped out of a tertiary course will be flagged for a certain number of years following their dropping out (to be determined by analysis of data) and their probability of returning in the following year calculated for each year. It is estimated that this probability of returning will diminish as years pass, and at a certain point the individual will lose their flag and be modelled with the rest of the workforce.

Alignment: As shown in Table 2, most of the education modules reviewed involve alignment of results. It is recommended that APPSIM's education module be aligned to targets based on administrative data from the Department of Education, Science and Training, or other data if required. A minimum level of detail for these targets could hopefully be the number of students in each type of education by age and sex. Further investigation of data will reveal the level of detail to which projections can be aligned. It would be ideal for other aspects of the education module, such as the modelling of HECS / HELP debt, to also be aligned to administrative data wherever possible.

Modelling student employment as a separate module: There was little information available for many of the models about how students' earnings were modelled while they were studying, probably indicating that did not form part of the education modules for most of the models. DYNAMOD and LifePaths are the only two education modules known to include the modelling of employment for students as separate to the models' labour force module. While there are arguments for why this was an appropriate way to model student earnings for DYNAMOD, further discussion with NATSEM staff who worked on the development of DYNAMOD indicates that this modelling was very complicated and perhaps should not be included in APPSIM's education module. There were also specific plans for the use of DYNAMOD at the time of its development that made the student earnings module necessary that do not apply to APPSIM. As discussed in Harding (2007), a broad lesson drawn from the overall review of dynamic microsimulation models is that in order to meet budgets and time frames it is most effective to first develop a simple working model that can be further developed if given time and funding. Thus it is recommended that a student earnings module is not included in APPSIM's education module at this stage.

Qualifications imputed: The models reviewed impute educational qualifications to varying levels of detail. However, generally they differentiate between secondary school, university and trade qualifications. The level of detail that can be modelled is governed by the purpose of the module within the whole model (how the qualifications will be used to model other things, and whether they are of interest in themselves), the availability of data and the appropriateness of attributing specific qualifications based on the content of the data.

A key function of APPSIM's education module is to inform the labour force module. The labour force module requires four levels of qualifications to be modelled: university, trade, year 12, and less than year 12 (Keegan 2007). This is therefore the minimum level of detail required to be modelled by the education module. Further review of data and analysis of its content will reveal whether a finer level of detail can and should be modelled. For example, DYNAMOD modelled the decision to complete year 11 and then year 12 separately, so that someone could have a highest qualification of year 10, 11 or 12. However if this level of detail is not necessary for the modelling of labour force status, or other life events, the level of complexity to model these two qualifications separately may outweigh the benefits of its inclusion. Alternatively, in order to meet targets of the correct number of students by age and sex it may be necessary to model the transitions between the two years separately, in which case the two qualifications could be attributed without any unnecessary complexity. Similarly at the other end of the qualification scale, small sample sizes could make the modelling of various post graduate qualifications problematic. If this level of detail is not required to inform other modules, and if the numbers are not of interest in themselves, simplification of the levels of qualifications may be appropriate.

Field of study: Most of the models reviewed do model the field of study undertaken by tertiary students. However, this modelling would be complex and possibly not of great additional value to the APPSIM model. As with detailed levels of qualification, the appropriateness of modelling the field of study depends on whether it is of interest to the model's users in itself, or required to inform other modules. There are arguments for and against modelling the field of study in APPSIM. It would require complex modelling and it is not yet clear whether it would be necessary, and a valuable use of time and budget. As people are increasingly changing career paths throughout their working lives, the influence of one's field of study on employment has become less clear. Field of study is not required by the labour force module. Alternatively, the number of students in various courses may be of intrinsic interest. Most importantly, field of study does impact on a student's HECS / HELP debt.

However the way in which this will be modelled, and what data are available is not yet clear. (This is discussed further below). While field of study does influence a student's debt, fees are *per unit,* not for the whole course, so field of study could not simply be used as a proxy for the level of debt held anyway. Perhaps instead an average debt could be allocated to students on the basis of other characteristics (to be explored when more is known about available data). Given these issues, at this stage it is recommended that field of study not be modelled by the APPSIM education module.

The use of 'macro' variables in the calculation of transition probabilities: DYNAMOD and MOSART include macro variables such as participation rates and relative wages in their education modules. This seems like an ideal inclusion in the modelling of educational trends. However, it would be complex as such factors are already implicit in administrative and survey data aggregates and the effect of such factors on what these aggregates *could* have been is not clear. The inclusion of a 'mic-mac' module in APPSIM - by which the 'micro' level life events of individuals are continually aligned against a 'macro' model modelling aggregates simultaneously - could possibly allow the impact of fluctuations at the macro level to be included in APPSIM's education modelling. However at this stage, in the interests of simplicity, it is recommended that macro variables not be included in the calculation of transition probabilities in APPSIM's 'stage one' education module.

3 Review of available data

The design of the module is also dictated by what data are available for use in modelling, and for validation. This section briefly outlines available data on educational stocks and flows. Appendix 1 to this paper contains some tables showing the distribution of Australians across various types of educational participation. As 2001 is the base year for APPSIM the review has focussed on data sources for that period, although data that is released at regular and frequent intervals is ideal.

3.1 Australian Census of Population and Housing, Household Sample File

Compiled by: Australian Bureau of Statistics (ABS)

Released: 2001, next release will be for the 2006 Census

The Census Household Sample File 2001 (HSF01) is a one per cent sample of the Australian population, containing 79,320 families and 188,013 persons (ABS, 2003).

As it is a one per cent sample, the weight used is 100 for each person. The HSF01 provides useful information on highest qualifications held, as well as the level and type of institution at which individuals are studying. Some information is also provided on when the highest qualification was completed in single years over the recent period, which could be useful for looking at transitions. There is also a variable stating whether the individual is studying full or part-time, presented in Figure 2.

The regular release and large sample size are advantages of using the HSF01. It will also be especially useful for the student labour market module as information on both educational and labour force participation is provided.

3.2 The Household, Income and Labour Dynamics in Australia (HILDA) Survey

Compiled by: The Melbourne Institute, University of Melbourne

Released: the fifth wave of the survey has been released in 2007 covering data from 2001 to 2005, with the next expected in 2008, covering data to 2006.

The HILDA survey is a longitudinal survey that was first conducted in 2001. The wave 1 panel consisted of 7682 households and 19,914 individuals. All individuals aged 15 and over are interviewed and give information on current and previous study to a detailed level. Information on labour force participation is also collected. Although the survey is a representative sample of the Australian population, when it is broken down by single year of age and gender, the unweighted sample in each type of education is quite small.

3.3 The Longitudinal Surveys of Australian Youth (LSAY)

Compiled by: Australian Council for Educational Research and Department of Education, Science and Training

Released: the fifth wave of the survey has been released in 2007 covering data from 2001 to 2005, with the next expected in 2008, covering data to 2006.

The LSAY conducts annual telephone interviews of several cohorts of Australian students, beginning when they are in mid secondary school until they are about 25. The samples are nationally representative. The oldest group in the project comprised people born in 1961; the youngest group comprises people born between 1 May 1987 and 30 April 1988. Information on the links between social characteristics, education and training, and employment can be drawn from the surveys. Issues investigated in the LSAY project include school achievement, school completion, participation in vocational and university education, gaining and maintaining employment, and household and family formation.

LSAY would be useful for 'flows' data, as it could provide information on 'who' is commencing or finishing what course, and as it is longitudinal it can provide insight into the paths people take.

3.4 Selected Higher Education Statistics

Compiled by: Department of Education, Science and Training (DEST)

Released: the most recently released data is from 2003, it is released annually

DEST releases administrative data on all students in Australian universities. Information is available on the number of students by age, gender, institution, citizenship, level and field of study, 'equity group' (for example rural, low socio-economic status, women in non-traditional area, indigenous). Notably there is a specific table on HECS / HELP status. There are also specific tables showing the numbers of commencing students, students new to higher education and completions.

3.5 Vocational and Technical Education Statistics

Compiled by: The National Centre for Vocational Education Research (NCVER)

Released: annually or quarterly, the most recent data is available for 2005 or 2006 depending on the table

A wealth of data is available on technical and vocational education. NCVER provides data under the broad categories of apprentices and trainees, students and courses and student outcomes. Thus, this data will be useful for modelling and /or validating numbers of students in technical and vocational education. The apprentices and trainees tables give the number of students by state, age (at commencement), qualification level, whether they are an existing worker, full or part-time status, expected duration of training and occupation.

4 Proposed approach to APPSIM

The review of the education modules of various dynamic microsimulation models and education data available in Australia presented here provide a foundation for proposing how education will be modelled in APPSIM. The aim of this proposed 'first cut' approach is to get a simple module working that will meet the broad requirements and purposes of APPSIM, which can be further developed given time and adequate budget. Further review and analysis of data will continue to inform the specifics of the education module, as will the iterative process of the module's development. The following section sets out a simplified proposal of what is required of APPSIM's education module, and how this will be achieved.

The design of the education module of a dynamic microsimulation model is dictated by the purpose of the module; what educational information is required; and how it will be used in other modules and output. On this basis there are three key points of information that need to be generated by the education model. These are:

1. Current study status (whether or not studying, full / part-time study and type of institution);

2. Highest qualification (to be updated as student completes a new course); and

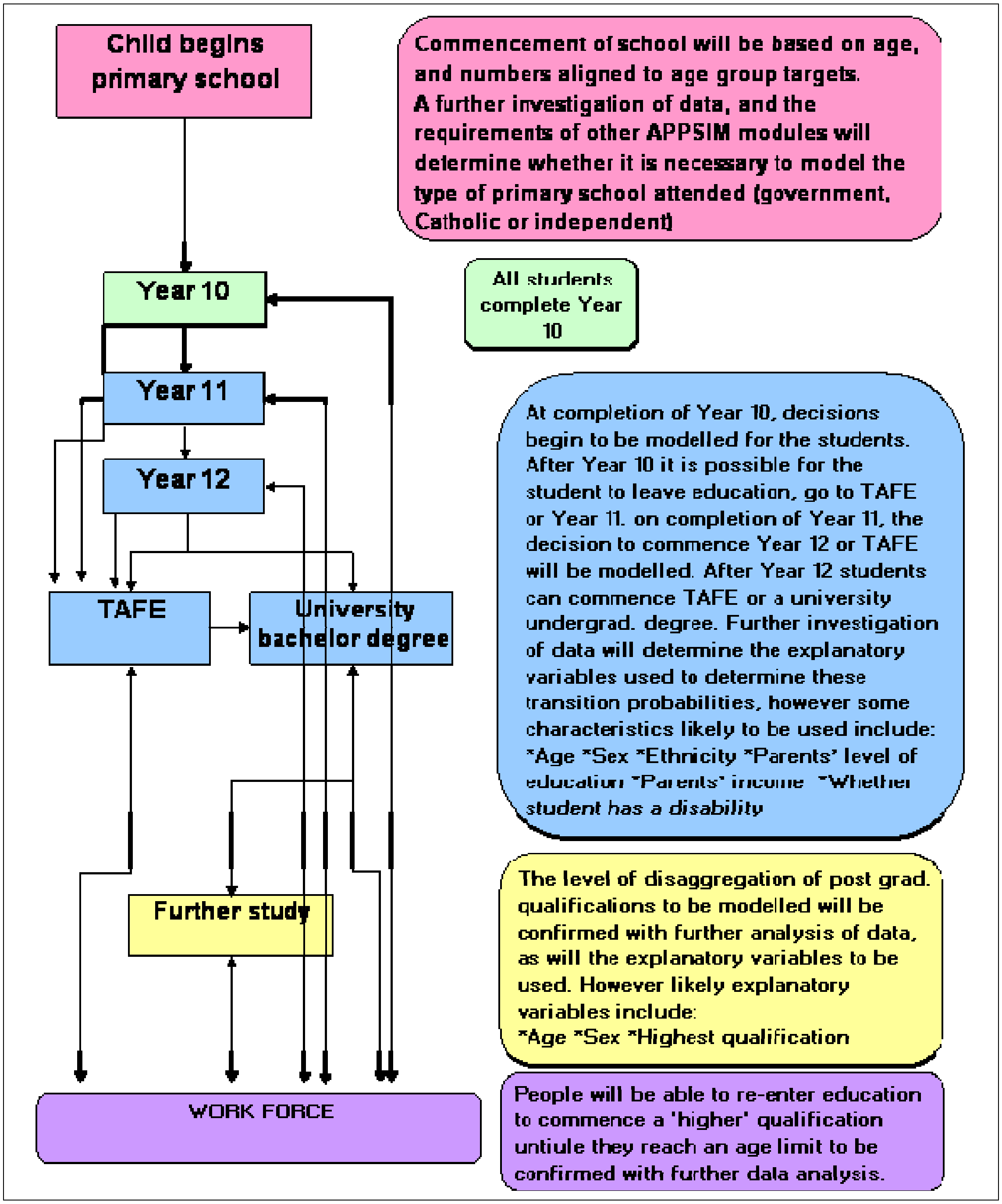
3. HECS / HELP 'band' and amount accumulated.

4.1 Current study status and highest qualification

The modelling of the first two points forms the major part of the education module. Based on the review of approaches to modelling education and available data discussed in this paper, as well as the requirements and aims of the APPSIM model, it is proposed that APPSIM's education module be similar to that of DYNAMOD, as shown in Figure 1, but with several changes. Figure 1 depicts the 'flow' of the DYNAMOD education module; the decisions made, who is at risk of each decision and the explanatory variables used to model each decision. This is the basic framework (detailed in Section 1 of this paper) that will be followed in APPSIM. The revised approach is shown in Figure 3.

Further data analysis will determine exactly which data sources will be used to model and align each transition through education. Regression based transition probabilities will be used to guide people from Year 10 through each of the possible paths shown in Figure 3, possibly based on the suggested explanatory variables shown. This process will designate for individuals both their current study status (including what type of institution they are attending and whether they are full or part time), and their current highest level of qualification.

Figure 3 **Proposed method for APPSIM's 'stage one' education module**



APPSIM's education module will also have the following differences to DYNAMOD:

- DYNAMOD used state of residence to model a child's commencement of primary school, the type of school they attended and secondary and tertiary educational transitions. In APPSIM the use of state in education modelling would be problematic as it would require the modelling of interstate migration for each individual throughout the course of their life. Further, the age at which children began school used to differ with state. This is to be phased out by 2010 so the minimum age for starting school will be the same for children in all states (see Atelier Learning Solutions and Access Economics 2006). Therefore state will not be used as an explanatory variable in the modelling of education in APPSIM. It is also recommended that the occupation of the potential student's parents not be used, as this will probably change throughout their life. In general further research will be conducted into what explanatory variables will be used for each transition.

- While education is generally based on calendar years, APPSIM is based on financial year periods. However this does not create a great problem as what is actually being modelled is the student's *transition within the previous period*, i.e. in the financial year 2001-02, this 15 year old commenced Year 10. Following the DYNAMOD method of all primary school students completing Year 10 before the possibility of dropping out or commencing other education is modelled, the proportion of the year spent in study would not need to be modelled, as before tertiary education it is not relevant for the purposes of the model. For those in tertiary education, part or full-time status will be modelled. Thus for the purposes of modelling income and HECS, the proportion of the year spent studying and working will be captured by this sub-module.

- It is worth noting that age regulations for starting school state a minimum age and that students will often begin a year older than this depending on what time of year they were born. Thus in order to match historical data, some students will need to begin at age five while others will begin at six. This distribution of an age group across various levels of school then continues throughout their schooling. Therefore rather than allocating a student transition into to a year level based on the age they turn in the period (i.e. at 16 a student completes Year 10), transitions will simply follow from Kindergarten to Year 10. Allocations will be validated against data on the distribution of people in each age group across various educational levels (similar to the tables in the Appendix) to make sure they capture the right number of say, 12 year old boys and girls in years 6 and 7. Further research will also confirm whether the assumption of all students completing Year 10 remains viable.

4.2 Modelling the Higher Education Contribution Scheme / Loan Program (HECS / HELP)

HECS / HELP is an Australian government scheme which allows students to defer their university course fees until their income exceeds a certain amount, after which a proportion of income is deducted until the loan is paid off. In 2005, HECS was replaced by HELP, although the scheme is often still colloquially referred to as HECS.

In 2001, around 79 per cent of university students had deferred their HECS liability - that is, they had not paid for their study up front and had incurred a HECS debt.

In order to construct to a simple model of HECS, the following information would be ideal:

- 'Who' is likely to defer their HECS debt? The frequency of deferred HECS debts by characteristics such as age, sex and any others available would be useful.
- How much debt is accumulated by who? The amount of debt accumulated per year by as detailed level of student characteristics as possible, hopefully eliminating the need for inclusion of 'field of study', as discussed above.

It is likely that data containing such information will be available from DEST and a further investigation of this data will determine how the debt will be modelled. Based on assumptions of what the data will contain, the simple model of HECS outlined below is proposed.

First, it would be necessary to flag which university students will defer their university fees (that is incur a HECS debt). If it is possible to obtain unit record data on university students and some of their characteristics, regression will be used to allocate a HECS debt to individuals. This is likely to be a high proportion and possibly needs to be aligned to the administrative data.

Second, the level of debt needs to be allocated to the individual. This is not a particularly simple thing to model, as HECS is accumulated for each unit a student does each semester, not simply at a broad course level. There are three 'bands' of HECS that apply to various field areas. This means that a single figure cannot be allocated to students when they complete a degree. It is recommended that a regression model be used to attribute an amount of debt to each student at the end of their course based on the length of their course at completion and other characteristics that are shown to impact on the level of HECS accumulated.

5 References

Atelier Learning Solutions and Access Economics. 2006, *Cost benefit analysis relating to the implementation of a common school starting age and associated nomenclature by January 1 2010 (Final Report),* Volume 1, Report for the Ministerial Council on Education, Employment, Training and Youth Affairs, Melbourne.

Australian Bureau of Statistics (ABS). 2001, *Census of Population and Housing Household Sample File, Confidentialised Unit Record File,* Australian Bureau of Statistics, Canberra.

ABS, 2003, *Technical Paper: Census of Population and Housing Household Sample File, Australia, 2001,* Catalogue No. 2037.0., Australian Bureau of Statistics, Canberra.

Favreault, M. and Smith, K. 2004, *A primer on the dynamic simulation of income model (DYNASIM3)*, Discussion Paper 02-04, Urban Institute, USA.

Flood, L., Jansson, F., Pettersson, T., Sundberg, O., and Westerberg, A. 2003, *SESIM III – a Swedish dynamic microsimulation model,* available at http://www.sesim.org/Documents/Handbook%20of%20SESIM.pdf

Fredriksen, D. 1998, *Projections of population, education, labour supply and public pension benefits: analyses with the dynamic microsimulation model MOSART*, Social and Economic Studies Paper 101, Statistics Norway, Olso and Kongsvinger.

Fredriksen, D. 2003, 'The MOSART Model – a Short Technical Documentation', Paper presented to the International Microsimulation Conference on Population, Ageing and Health: Modelling Our Future, Canberra, December. Available at http://www.natsem.canberra.edu.au/conference2003/papers/pdf/stoelen_nils-2.pdf

Gupta, A. and Harding, A. (eds), 2007, *International Symposia in Economic Theory and Econometrics*, North Holland, Amsterdam

Hardin, A. 1994, 'Modelling education in DYNAMOD', Dynamic Technical Paper No. 5, NATSEM, University of Canberra, Canberra.

Harding, A. 2007, Challenges and opportunities of dynamic microsimulation modelling, Plenary paper presented at the 1st General Conference of the International Microsimulation Association, Vienna, August, 2007.

Harding, A. and Gupta, A. 2007, *Modelling our Future: Population Ageing, Social Security and Taxation*, North-Holland, Amsterdam.

Keegan, M. 2007, Modelling the workers of tomorrow: the APPSIM dynamic microsimulation model, Paper presented to the HILDA Survey Research Conference, 19-20 July.

Morrison, R. and Dussault, B. 2000, *Overview of DYNACAN, Canada*, Available at http://www.actuaries.org/CTTEES_SOCSEC/Documents/dynacan.pdf

Pylkkänen, E. 2000, 'Modelling wages and hours of work', paper presented at the 6th Nordic Seminar on Micro Simulation Models, Copenhagen, Denmark, 8-9 June.

Robinson, M. and Bækgaard, H. 2002, *Modelling students in DYNAMOD-2*, NATSEM Technical Paper no. 23, National Centre for Social and Economic Modelling, University of Canberra, Canberra.

Statistics Canada, *The lifepaths microsimulation model: an overview*, last accessed March 2007, Available at http://www.ssc.uwo.ca/sociology/popchange/LifePathsOverview_E.pdf

Scott, A. and Zaidi, A. 2004, *Education and labour market predictors in the SAGE dynamic microsimulation model*, SAGE Technical Note No. 9, London School of Economics, London.

Appendix 1: Census 2001 Tables

Table 1: **Number and percentage of females participating in education by age and education type**

Age years	Pre-school	Government Primary	Non-Government Primary	Government Secondary	Non-Government Secondary	TAFE	University	Other, NS,NA, OS Visitor*	Total
3	26,300	.	.	.	.	.	.	99,200	126,300
%	*21*	.	.	.	.	.	.	*79*	
4	65,800	.	.	.	.	.	.	49,100	121,300
%	*54*	.	.	.	.	.	.	*40*	
5	38,300	49,200	23,600	.	.	.	.	15,100	126,400
%	*30*	*39*	*19*	.	.	.	.	*12*	
6	.	77,900	34,500	.	.	.	.	12,000	125,000
%	.	*62*	*28*	.	.	.	.	*10*	
7	.	79,300	34,700	.	.	.	.	10,200	124,900
%	.	*63*	*28*	.	.	.	.	*8*	
8	.	83,300	30,700	.	.	.	.	12,000	126,700
%	.	*66*	*24*	.	.	.	.	*9*	
9	.	88,600	35,800	.	.	.	.	11,100	136,300
%	.	*65*	*26*	.	.	.	.	*8*	
10	.	89,500	33,700	.	.	.	.	13,100	136,700
%	.	*65*	*25*	.	.	.	.	*10*	
11	.	81,100	33,200	.	.	.	.	8,800	128,900
%	.	*63*	*26*	.	.	.	.	*7*	
12	.	42,600	18,000	37,300	24,900	.	.	6,700	130,600
%	.	*33*	*14*	*29*	*19*	.	.	*5*	
13	.	4,800	.	69,300	44,200	.	.	5,400	126,400
%	.	*4*	.	*55*	*35*	.	.	*4*	
14	.	.	.	72,900	43,500	.	.	7,800	129,200
%	.	.	.	*56*	*34*	.	.	*6*	
15	.	.	.	66,300	39,800	.	.	12,800	123,500
%	.	.	.	*54*	*32*	.	.	*10*	
16	.	.	.	64,100	39,000	4,800	.	19,500	132,200
%	.	.	.	*48*	*30*	*4*	.	*15*	
17	.	.	.	55,600	32,400	7,500	5,800	30,200	134,200
%	.	.	.	*41*	*24*	*6*	*4*	*23*	

Table 1 continued

Age years	Pre-school	Government Primary	Non-Government Primary	Government Secondary	Non-Government Secondary	TAFE	University	Other, NS,NA, OS Visitor*	Total
18	.	.	.	12,300	8,200	16,900	33,100	50,700	126,600
%	.	.	.	*10*	*6*	*13*	*26*	*40*	
19	.	.	.	.	.	20,600	43,300	60,200	132,400
%	.	.	.	.	.	*16*	*33*	*45*	
20	.	.	.	.	.	11,100	44,800	62,300	123,900
%	.	.	.	.	.	*9*	*36*	*50*	
21	.	.	.	.	.	9,700	36,700	76,400	124,200
%	.	.	.	.	.	*8*	*30*	*62*	
22	.	.	.	.	.	8,200	26,200	77,200	118,000
%	.	.	.	.	.	*7*	*22*	*65*	
23	.	.	.	.	.	7,500	21,900	98,200	130,200
%	.	.	.	.	.	*6*	*17*	*75*	
24	.	.	.	.	.	6,600	15,700	94,300	123,000
%	.	.	.	.	.	*5*	*13*	*79*	
25–29	.	.	.	.	.	30,900	53,500	599,700	684,400
%	.	.	.	.	.	*5*	*8*	*88*	
30–34	.	.	.	.	.	28,200	36,600	658,200	723,300
%	.	.	.	.	.	*4*	*5*	*91*	
35–39	.	.	.	.	.	28,200	28,700	667,700	725,100
%	.	.	.	.	.	*4*	*4*	*92*	
40–44	.	.	.	.	.	26,000	22,200	683,100	732,300
%	.	.	.	.	.	*4*	*3*	*93*	
45–49	.	.	.	.	.	20,800	17,900	631,700	671,100
%	.	.	.	.	.	*3*	*3*	*94*	
50–54	.	.	.	.	.	13,000	9,800	608,700	632,000
%	.	.	.	.	.	*2*	*2*	*96*	
55–59	.	.	.	.	.	5,800	4,200	474,400	487,800
%	.	.	.	.	.	*1*	*1*	*97*	
60–64	.	.	.	.	.	.	.	393,300	399,200
%	.	.	.	.	.	.	.	*99*	
65–69	.	.	.	.	.	.	.	334,600	341,100
%	.	.	.	.	.	.	.	*98*	
70–74	.	.	.	.	.	.	.	321,700	326,800
%	.	.	.	.	.	.	.	*98*	
75–79	.	.	.	.	.	.	.	277,000	279,500
%	.	.	.	.	.	.	.	*99*	

Table 1 continued

Age years	Pre-school	Government Primary	Non-Government Primary	Government Secondary	Non-Government Secondary	TAFE	University	Other, NS,NA, OS Visitor*	Total
80–84	.	.	.	.	.	.	.	197,500	199,100
%	.	.	.	.	.	.	.	*99*	
85 and over	.	.	.	.	.	.	.	182,000	182,900
%	.	.	.	.	.	.	.	*100*	
Total	130,400	601,600	249,600	388,200	236,800	252,100	404,200	7,278,000	9,540,900
%	*1*	*6*	*3*	*4*	*2*	*3*	*4*	*76*	*100*

Source: ABS Census 2001 Household Sample File CURF.

Note**:** Cells referring to small samples have been removed.

*This column combines 'Other', 'Not Stated', 'Not Applicable' and 'Overseas Visitor'

Table 2 **Number and percentage of males participating in education by age and education type**

Age years	Pre-school	Government Primary	Non-Government Primary	Government Secondary	Non-Government Secondary	TAFE	University	Other, NS,NA, OS Visitor*	Total
3	27,100	.	.	.	.	.	.	97,100	124,500
%	*22*	.	.	.	.	.	.	*78*	
4	69,400	4,400	.	.	.	.	.	53,200	130,700
%	*53*	*3*	.	.	.	.	.	*41*	
5	46,000	49,800	19,900	.	.	.	.	17,000	133,100
%	*35*	*37*	*15*	.	.	.	.	*13*	
6	.	83,800	28,900	.	.	.	.	14,800	128,300
%	.	*65*	*23*	.	.	.	.	*12*	
7	.	85,000	34,900	.	.	.	.	12,600	133,300
%	.	*64*	*26*	.	.	.	.	*9*	
8	.	84,900	36,700	.	.	.	.	11,500	134,000
%	.	*63*	*27*	.	.	.	.	*9*	
9	.	*88,000*	*33,900*	.	.	.	.	11,800	134,600
%	.	65	25	.	.	.	.	*9*	
10	.	88,900	34,400	.	.	.	.	13,300	137,400
%	.	*65*	*25*	.	.	.	.	*10*	
11	.	88,100	37,500	.	.	.	.	8,700	139,200
%	.	*63*	*27*	.	.	.	.	*6*	
12	.	48,300	17,100	35,500	22,600	.	.	9,600	134,000
%	.	*36*	*13*	*26*	*17*	.	.	*7*	
13	.	8,700	.	71,900	39,300	.	.	6,000	129,700
%	.	*7*	.	*55*	*30*	.	.	*5*	
14	.	.	.	76,500	41,300	.	.	8,300	132,700
%	.	.	.	*58*	*31*	.	.	*6*	
15	.	.	.	78,000	41,200	.	.	18,600	141,000
%	.	.	.	*55*	*29*	.	.	*13*	
16	.	.	.	66,000	41,600	7,700	.	23,500	142,000
%	.	.	.	*46*	*29*	*5*	.	*17*	
17	.	.	.	48,600	30,000	10,700	.	38,800	134,800
%	.	.	.	*36*	*22*	*8*	.	*29*	
18	.	.	.	13,300	8,300	25,000	23,800	61,000	135,600
%	.	.	.	*10*	*6*	*18*	*18*	*45*	
19	.	.	.	.	.	24,500	32,200	75,200	137,900
%	.	.	.	.	.	*18*	*23*	*55*	

Table 2 continued

Age years	Pre-school	Government Primary	Non-Government Primary	Government Secondary	Non-Government Secondary	TAFE	University	Other, NS,NA, OS Visitor*	Total
20	.	.	.	.	.	19,200	31,000	80,500	135,200
%	.	.	.	.	.	*14*	*23*	*60*	
21	.	.	.	.	.	14,500	26,800	79,900	127,500
%	.	.	.	.	.	*11*	*21*	*63*	
22	.	.	.	.	.	8,900	22,300	87,800	122,600
%	.	.	.	.	.	*7*	*18*	*72*	
23	.	.	.	.	.	7,800	16,600	93,400	123,600
%	.	.	.	.	.	*6*	*13*	*76*	
24	.	.	.	.	.	5,700	14,000	100,000	125,400
%	.	.	.	.	.	*5*	*11*	*80*	
25–29	.	.	.	.	.	28,800	47,000	586,000	662,100
%	.	.	.	.	.	*4*	*7*	*89*	
30–34	.	.	.	.	.	24,100	32,700	639,500	696,700
%	.	.	.	.	.	*3*	*5*	*92*	
35–39	.	.	.	.	.	18,700	22,200	662,700	703,900
%	.	.	.	.	.	*3*	*3*	*94*	
40–44	.	.	.	.	.	14,200	17,500	664,600	696,700
%	.	.	.	.	.	*2*	*3*	*95*	
45–49	.	.	.	.	.	12,300	10,500	633,100	660,100
%	.	.	.	.	.	*2*	*2*	*96*	
50–54	.	.	.	.	.	8,500	7,000	604,800	623,500
%	.	.	.	.	.	*1*	*1*	*97*	
55–59	.	.	.	.	.	4,700	.	483,200	493,200
%	.	.	.	.	.	*1*	.	*98*	
60–64	.	.	.	.	.	.	.	393,400	400,700
%	.	.	.	.	.	.	.	*98*	
65–69	.	.	.	.		.	.	323,500	328,200
%	.	.	.	.		.	.	*99*	
70–74	.	.	.	.	.	.	.	282,800	286,200
%	.	.	.	.	.	.	.	*99*	
75–79	.	.	.	.	.	.	.	212,000	214,000
%	.	.	.	.	.	.	.	*99*	
80–84	.	.	.	.	.	.	.	116,900	117,700
%	.	.	.	.	.	.	.	*99*	

Table 2 continued

Age years	Pre-school	Government Primary	Non-Government Primary	Government Secondary	Non-Government Secondary	TAFE	University	Other, NS,NA, OS Visitor*	Total
85 and over	.	.	.	.	.	.	.	82,300	83,000
%	.	.	.	.	.	.	.	*99*	
Total	142,500	632,900	249,900	396,200	227,700	240,600	313,000	7,057,600	9,260,400
%	*2*	*7*	*3*	*4*	*2*	*3*	*3*	*76*	*100*

Source: ABS Census 2001 Household Sample File CURF.

Note**:** Cells referring to small samples have been removed.

*This column combines 'Other', 'Not Stated', 'Not Applicable' and 'Overseas Visitor'