COMMONWEALTH GRANTS COMMISSION

DISCUSSION PAPER CGC 2003/03

EDUCATION — ANALYSIS OF POST-COMPULSORY ENROLMENTS

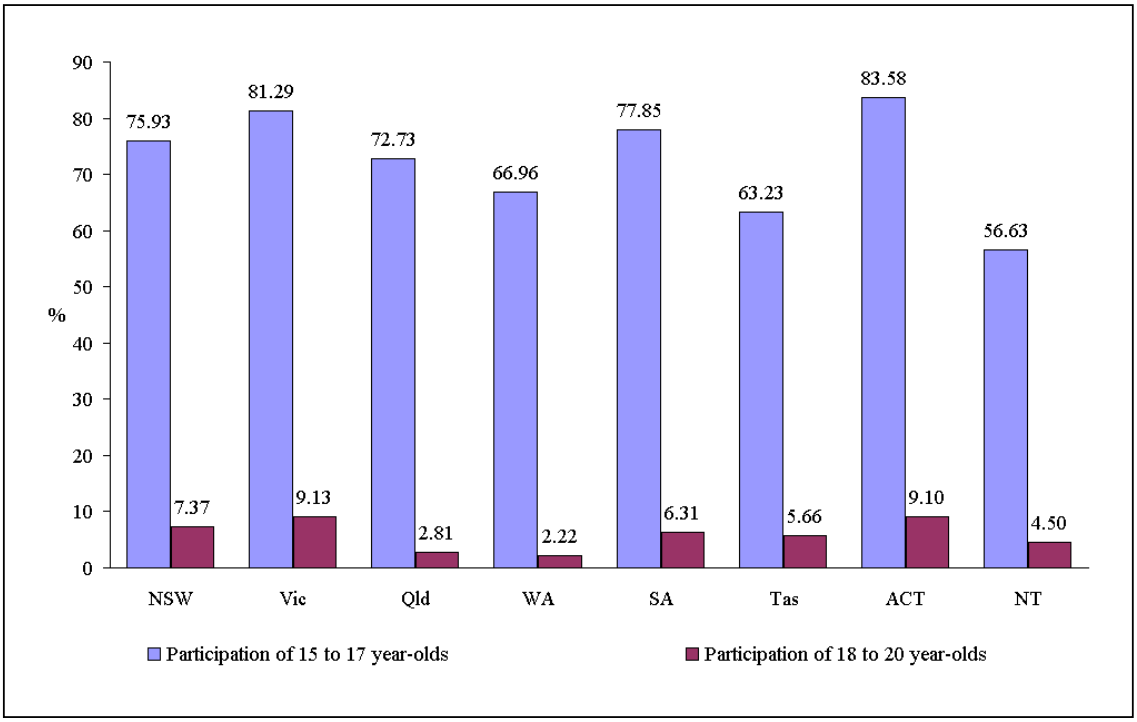
JUNE 2003

CONTENTS

INTRODUCTION

1. This Discussion Paper outlines the work that the Commission staff has undertaken to analyse the influences on post-compulsory enrolments. This work involved exploring options to measure the impact of policy and non-policy influences on post compulsory enrolments in the 2004 Review assessments. The actual participation rates of cohorts central to this work are in Figure 1.

Figure 1 COMPARISON OF SCHOOL PARTICIPATION RATES OF 15-17 YEAR OLDS AND 18 TO 20 YEAR-OLDS, BASED ON 2001 ABS CENSUS



Major issues for the 2004 Review

2. The relative differences in post-compulsory enrolment rates across the States are thought to be due to the interplay of policy and non-policy effects. The challenge for the Commission when calculating standardised post-compulsory enrolments is to design an analytical model which accurately measures:

(i) average levels of enrolments across the States; and

(ii) any variations around this average that reflect disabilities[1].

3. The submissions made by the States, the discussions with them and the large body of literature reviewed all show notable divergence of opinion over:

(i) a suitable measure of post-compulsory enrolments; and

(ii) the method of separating policy influences from non-policy influences.

4. In the 1999 Review, the Commission based its measurement of post-compulsory enrolments on an analysis of participation rate data. It used a regression analysis of the data on actual participation and corresponding information on socio-economic status and location to estimate the impact of these characteristics on post-compulsory enrolments. The intercept of the regression model was treated as the measure of average enrolments across the States. The disabilities were estimated using the regression output.

5. Participation rate data were used as the basis of the analysis in the 1999 Review because:

(i) the Commission research showed that age participation rates were more uniform across States than grade retention rates[2];

(ii) relevant data were more accessible at a small area level, such as the SLA, enabling the use of rich data describing socio-economic status[3]; and

(iii) the analysis could be done using data for a single year.

6. The alternatives proposed by the States for this review include retention rates and completion rates.

7. Victoria put forward a retention rate model using time series data which had been carried out by Dr Stephen Lamb. The analysis was based on the Longitudinal Surveys of Australian Youth (LSAY). LSAY data comprises a series of past and continuing cohort surveys of Australian youth, with data covering a number of issues including education. The original survey sample selected cohorts, each of over 13 000 Year 9 students in 1995 and 1998. Information with regard to each of these sample cohorts has been updated annually.

8. In Discussion Paper 2002/28, staff said that issues affecting the use of these data for Commission purposes may include the adequacy of State-specific representation of metropolitan, regional and rural populations in the samples, the representation of Indigenous people, the effects of sample attrition and the sample weighting procedure to

---

1 These are influences on the average service which are beyond the control of States.

2 See for example the *Report on General Revenue Grant Relativities 1993,* Working Papers 1987-88 to 1991-92, Commonwealth Grants Commission, May 1993, page 100.

3 See page 8, CGC 2002/28.

represent the population of Year 9 students. Staff said its preference was to undertake the analysis of participation rate data rather than explore further the analysis of retention using the LSAY data. The paper said that the use of Census data would give the capacity to accurately assess the influence of factors such as Indigeneity and location.

9. Following the November staff level conference Victoria also presented analysis using 2001 Census data for secondary school completion rates. This work had also been carried out by Dr Lamb. In summary, the paper proposed a regression model using school completion rates for 19 and 20 year olds to analyse influences on the level of post-compulsory enrolments. The dependent variable was expressed as a per capita rate and was based on the numbers of 19 and 20 year olds who reported in the 2001 Census that they had completed secondary education. The independent variables used in the model included an education and occupation based SEIFA and a mix of policy and non-policy variables. With this combination of variables, Lamb estimated a model that explained influences on post-compulsory enrolments to a significantly larger degree than had been possible through the Commission's 1999 Review approach.

10. Attachment D provides a summary of the comments provided by the States in rejoinder submissions on the issues relevant to the analysis reported in this paper. In general, the States did not support use of retention rates as the basis for investigating the affects of non-policy influences on post-compulsory enrolments. Some States supported the use of average participation rates derived from Census data to calculate standardised enrolments as an alternative to estimating standardised enrolments using regression analysis.

Outline of the work

11. The work undertaken since the November conference involved using regression techniques to investigate policy and non-policy affects on post–compulsory enrolments. More specifically, it involved:

(i) specifying a regression model based on participation rates while taking into account the key concerns raised about the 1999 Review model — particularly, the measurement and modelling of post-compulsory schooling and data issues;

(ii) running the completion rate model proposed by Lamb, replacing SEIFA with more specific variables, adjusting for the effects of mobility and adding policy variables; and

(iii) comparing the outcome from the participation rate model with those from the models based on completion rate data proposed by Lamb and Australian average participation rates.

THE PARTICIPATION RATE MODEL

12. A regression model using participation rates to estimate standard and standardised rates of post-compulsory enrolments has been used to address the issues raised by States and those identified in staff research. The model has benefited from the work done by Victoria and the ACT and other available empirical research[4]. The modifications we have made to the 1999 Review model include:

(i) replacement of SEIFA, an index designed by ABS to capture a select set of socio-economic variables, with a range of independent variables to more transparently explain the variability of post-compulsory enrolments across States; and

(ii) inclusion of specific variables to represent State policy and other effects to increase the capacity of the model to explain State differentials in post-compulsory enrolments.

13. We have also taken steps:

(i) to further increase the predictability of the model[5]; and

(ii) to ensure that the positive and/or negative effects of the policy and non-policy influences that were considered in the model accorded with prior expectations about them[6].

14. The relationship between post-compulsory participation rates and the various influences on it can be conceptually modelled as shown below.

$$PR_i = f\ (SEV_i, SP_{ij}, O_i), \quad i = 1, 2, \ldots.n$$

where PR_i is the participation rate of State i and is a continuous variable;

SEV_i are non-policy variables;

SP_{ij} are State policy variables (j = 1, 2, ….8); and

O_i is the error term.

15. The model is diagrammatically presented below:

---

4 See, for instance, Roussel and Murphy (2000) and Jones (2002). The paper by Roger Jones is entitled '*Educational Participation and Outcomes by Geographic Location',* (Australian Council of Education Research, Victoria, June 2002).

5 As given by the 'goodness of fit of the regression model' or its adjusted R^2 and the statistical significance of the model coefficients.

6 A number of diagnostic criteria were used for judging the practical validity of the model. These included checking whether the regression coefficients accorded with a priori assumptions about their behaviour.

Dependent Variable		Intercept		Independent Variables				
Participation Rate	=	Standard Participation Rate	+	Non-policy Variables	+	Policy Variables	+	Error Term

Data used in the model

16. The applicability of the proposed model was tested using ABS 2001 Census data at the statistical local area (SLA) level for 15 to 20 year olds and data from various published sources.

17. The ABS Census data used in the model were at the SLA level. These were available to us at two levels of aggregation[7]. The cross-tabulated data that we requested from the ABS were relatively less aggregated than that available from the 2001 CDATA.

18. The cross-tabulated data were for families of 15 to 20 year-olds participating in post-compulsory schooling. This data enabled us to identify and study the influence of family income, the availability of non-government schools in a SLA, ethnicity and Indigenity on school participants of any age from 15 to 20 years. The influence on participation of labour market conditions and of remoteness were modelled for the whole population in a SLA using CDATA.

19. We checked the 2001 Census cross-tabulated data against the other published source of school enrolment data — *Schools Australia* (Catalogue number: 4221.0). Attachment A provides a comparison of the data obtained from the ABS in cross-tabulated form and that published in the June 2001 issue of *Schools Australia*. There are some differences between the two data sources. Reasons for these differences include the following:

(i) Differing modes of enumeration — Census data are recorded by the respondents and as a result their perceptions and beliefs are reflected in the data. The *Schools Australia* data on the other hand are derived from State administrative records compiled according to consistent rules.

(ii) Differing definitions — A number of differences exist in the definition of data items used in the *Schools Australia* publication and the definition of data items in the Census. For example, *Schools Australia* defined a 'student' as a person who is not absent from school for a total period of four continuous weeks. Perceptions of a respondent at the last Census were not necessarily so rigorous[8].

20. We concluded that the Census data provided a reasonable basis for this analysis.

---

[7] ABS aggregates Census data to prevent identification of the characteristics of individual households.

[8] Personal communications with the ABS in April 2003.

21. We found considerable variability in the population sizes of SLAs[9]. For instance, 120 SLAs have been assigned by the ABS to the ACT which is a single administrative unit. This was done to facilitate data reporting. But population numbers in SLAs in the ACT are significantly fewer than the average population size found in comparable metropolitan areas in most other States. The implication for this present analysis is that the ACT (and other areas where the ABS has adopted this practice such as the Brisbane metropolitan area) has a much higher number of observations or entities in the modelling (in our model an observation was one SLA) than it would have if observations were apportioned according to population shares. We have applied a population weight to each independent variable used in the analysis to remove any bias that might be present due to the uneven population size of SLAs.

Elements of the participation rate model

22. Although persons aged 15 to 20 participate in post-compulsory schooling, our modelling research was mainly confined to 15 to 17 year-olds. Therefore in our primary participation rate model we used data for the participation rates of 15 to 17 year-olds in each SLA as the dependent variable. Various policy and non-policy influences that caused variability in the participation rate across SLAs were used as the independent variables.

23. In the secondary model, the data for the participation rates of 18 to 20 year-olds were used as the dependent variable – but we used exactly the same number and type of independent variables as in the primary model.

24. ***The dependent variable of the primary model.*** In the primary model, the dependent variable was calculated by taking the percentage of persons aged 15 to 17 in an SLA enrolled in post-compulsory education relative to the total population of this age group in that SLA. The reasons for using 15 to 17 year-olds in the primary modal are given below:

(i) Post-compulsory years of schooling usually commence at the age of about 15 years. Due to State differences in the age of school commencement and duration of stay in compulsory schooling, the age range 15 to 17 years can be considered as the important period for making decisions about participation.

(ii) Empirical evidence showed that the highest rates of participation were in the 15 to 17 years group and the rates of participation dropped markedly in the older age groups (see Figure 1 above).

---

[9] Out of a total of 1380 SLAs, only 1350 were used in the analysis. SLAs that did not physically exist and external territories were excluded. These were: (a) undefined (city/State), (b) no usual address and (c) external territories.

25. The dependent variable as used in the model is shown below:

$$\text{Participation Rate} = \frac{\text{Total persons 15 to 17 year-olds enrolled in post-compulsory schooling in a SLA}}{\text{Total persons aged 15 to 17 years in the SLA}} * 100$$

26. ***Independent variables.*** The independent influence of six non-policy and four policy variables on rates of participation at SLA levels were considered in the model.

27. ***Non-policy variables.*** These directly or indirectly reflect the socio-demographic characteristics of families on the participation of 15 to 17 year-olds in post-compulsory schooling. Brief descriptions of these variables are in Table 1. Detailed discussion of these variables is in Attachment B.

Table 1 NON-POLICY VARIABLES USED IN THE ANALYSIS

	Description
Family income:	A measure of the influence of the socio-economic status of families
Ethnicity:	A measure of the influence of being a person of ethnic origin.
Indigeneity:	A measure of the influence of being of Aboriginal and/or Torres Strait Islander background.
Remoteness:	A measure of the influence of location.
Full-time employment:	A measure of the influence of local labour market conditions on participation.
School sector	A measure of the influence of the availability of non-government schools.

28. ***Policy variables.*** We included four variables to account for State-specific policies affecting (a) school type, (b) age of commencement of post-compulsory schooling, (c) part-time studies and (d) participation in vocational education and training (VET). Brief descriptions of these variables are in Table 2. A detailed description of the reasons for selecting these variables and the relevant empirical literature is in Attachment B.

Table 2 POLICY VARIABLES USED IN THE MODEL

	Description
Age of commencement of post-compulsory schooling	A measure of the influence of policies affecting the age of commencement of post-compulsory schooling.
Proportion of part-time students	A measure of the influence of policies affecting the availability of part-time schooling.
School type	A measure of the influence of policies affecting the availability of secondary colleges.
Outside VET	A measure of the influence of policies affecting the general availability of vocational education and training outside the school system.

29. ***State-dummy variables.*** State effects arise due to the differential effects of State policies. Interactions between State policy and non-policy influences may also lead to significant State effects. It is difficult to identify and quantify these effects. Therefore in our model the State influences were represented using a set of binary dummy variables. These State dummies were intended to account for:

(i) any policy influences not captured by the four variables noted above; and

(ii) the particular circumstances in each State that affect the implementation of those State policies.

Results and interpretation

30. ***Model performance.*** Table 3 shows the results obtained for each of four different model specifications:

(i) Model 1 regresses participation rates on non-policy variables and includes an intercept;

(ii) Model 2 is the same as Model 1 but without an intercept;

(iii) Model 3 regresses participation rates on non-policy and policy variables and includes State dummies and an intercept;

(iv) Model 4 is the same as Model 3, but without an intercept.

31. Model 3, with intercept and policy variables, is our preferred model for predicting participation rates. It explained 63.9 per cent of the variability in participation. By comparison with Model 1, it can be seen that the introduction of policy variables increased the explanatory power of the model by 8 percentage points. Models 2 and 4 are included here to allow direct comparison with Lamb's models. We recognise, however, that models such as these which do not include an intercept or (in the case of Model 4) a full set of dummy variables[10] can give misleading results.

32. Numerous trials that we carried out using various transformations of variables as well as corrections for outliers failed to produce results that are distinctly different from those reported here. Therefore it can be inferred that the proportion of variance not explained by our analysis (about 36 per cent in Model 3) represents both policy and non-policy influences that have not been adequately measured. One important omission that we are aware of is the influence of State policies regarding the provision of vocational education and training in schools (VET in schools). Information available to us suggests that VET in schools is a policy variable of high significance relative to the other policy variables already included in the model. We think its inclusion has the potential to increase the influence of policy variables in the model, reduce the size of the error term and possibly change the coefficients for the non-policy variables.

---

[10] The model contains dummy variables for the States but excludes a variable for New South Wales.

33. Staff are taking steps to obtain within-State data for VET in schools. These data will enable inclusion of a meaningful measure of this policy influence in a future version of the model which will be produced after receipt of State comments on this present work.

34. With regard to the unexplained variability of the model there are other less tangible influences that may impact on decisions to participate in post-compulsory education. These include the level of engagement in school life, academic self concept, educational aspirations and parental expectations and the like. These are 'often neglected in models of participation in education' (Marks et al, 2000)[11]. It will be difficult to achieve accurate measurement of these influences in the type of model we have used here. Unit record data that included attitudinal variables would be required to do justice to these influences.

35. As noted above, the models without intercepts shown in the table enabled us to make some comparison of the results of our preferred model with those obtained in Dr Lamb's model — discussed later in the paper. The explanatory power (adjusted R^2) of the 'non-intercept' models appears considerably higher compared to those with intercepts. This is misleading as the adjusted R^2 values are not comparable when an intercept is not included[12].

36. We used the models with intercepts to predict standardised post-compulsory enrolments. This is because their results are less likely to be misleading than models without intercepts and they are intuitively more attractive. They calculate an Australian average participation rate which changes as State non-policy and policy variables are included. A model without an intercept forces participation rates to through the origin and allocates all variation to State variables. This takes no account of any underlying effects common to all States.

37. ***Influence of non-policy variables.*** In line with the findings of other researchers, our models showed that a substantial portion of the variability in participation was due to non-policy influences.

38. As expected, the models 1 to 4 showed that Indigeneity had a statistically significant negative influence on participation in post-compulsory education. Similarly all these models showed that ethnicity was positively and significantly correlated with participation rates. These observations accorded with our expectations.

39. Most models showed that the level of enrolments in non-government schools contributed significantly to participation in post-compulsory education — an expected result. Furthermore, all the models showed that family income had a positive and significant impact on participation. Available empirical evidence suggests that the choice

---

[11] Marks, N. G., Fleming, N., Long, M. and McMillan, J. (2000). Patterns of *Participation in Year 12 and Higher Education in Australia: Trends and Issues,* Research Report 17, Australian Council for Educational Research, Camberwell, Victoria, page vi.

[12] See for example http://pages.stern.nyu.edu/~adamodar/New_Home_Page/datafile/MReg97.html); and other university web sites like: http://www.biostat.wustl.edu/mailinglists/s-news/200007/msg00105.html).

of school sector is influenced by family income[13]. Nevertheless statistical tests showed that their level of interdependence[14] did not affect the model outcome.

Table 3 VALUES OF THE COEFFICIENTS FOR DIFFERENT MODELS USED IN THE ANALYSIS

Variable	Coefficients (Model 1)	Coefficients (Model 2)	Coefficients (Model 3)	Coefficients (Model 4)
Non-policy				
Constant	46.853		40.155	
Indigenity	-0.498**	-0.158	-0.370**	-0.341**
Ethnicity	0.034**	0.082**	0.021**	0.022**
School sector	0.079**	0.245**	0.072**	0.082**
Family income	0.050**	0.098**	0.049**	0.052**
Remoteness	-2.268	-0.489	-7.139**	-7.210**
Full-time employment	-0.535**		-0.310**	0.289**
Policy				
Secondary colleges			0.111*	0.079
Age of commencement			8.276	45.447**
Outside VET			-0.155**	-0.151**
Proportion of part-time students			-0.756**	-0.706**
VIC			2.368**	2.497**
QLD			6.620	43.896**
WA			1.984	39.162**
SA			9.905	47.169**
TAS			-9.164**	-7.688**
NT			-2.217	-2.451
ACT			3.099*	3.196*
Adjusted R^2	**0.559**	**0.988**	**0.639**	**0.993**

Notes: (a) ** coefficients with a t-statistic of 2.576 or above. These coefficients are significant at the 1% level of probability. * coefficients with a t-statistic of 1.960 or above. These coefficients are significant at the 5% level of probability. Other variables were significant at the 10% level.

Note that the significance of the influence of an independent variable on the dependent variable is assessed by the size of the t-statistic and its probability. The lower the probability, the more significant is the influence of the variable. A level of probability of 10 per cent or lower is a good indicator of the importance of an independent variable.

Models 3 and 4 were standardised on New South Wales by not including the dummy variable for NSW in them.

[13] See for instance Mukherjee (1999) *Socio-economic Status and School System Enrolments*, Australian Centre for Equity through Education, http://members.ozemail.com.au/~devm/SES.HTML.

[14] Tests for multi-colinearity were carried out to make this assertion.

40. As per empirical evidence, it was expected that the availability of full-time employment for persons aged 15 to 20 in a SLA would have a negative influence on post-compulsory enrolments[15]. The model confirmed this expectation.

41. Models 1, 3 and 4 showed that the participation of 15 to 17 year-olds in post-compulsory education was significantly reduced by remoteness. As noted in Attachment B, this result was consistent with the findings of other researchers.

42. ***Influence of policy influences.*** The policy variables included in the model on the whole performed satisfactorily. The variable for part-time students showed consistent influence on participation. As shown in Models 3 and 4, this influence was negative and highly significant.

43. The age of commencement in post-compulsory education showed a significant positive influence — especially in States where the school starting age was around 6 years compared to other States where the starting age was 5 years. As evident from the negative sign of the coefficient and its very high statistical significance, the variable for outside VET performed satisfactorily in both the models with policy variables.

44. It is more meaningful[16] to compare the State dummy variables within a model than between 'intercept and a non-intercept models'. Variations in State variable coefficients between models are magnified by the presence or absence of an intercept. The variations within the models on the other hand reflect the relative strength that each State has in augmenting socio-demographic influences towards participation.

45. The policy variables increased the explanatory power of intercept models by 8.0 per cent (compare Model 3 with Model 1). The corresponding increase in explanatory power for non-intercept models was only 0.5 per cent. These results are not comparable because they are a result of altering the natural geometric properties of the regression line by altering one or more of the regression coefficients[17].

46. Our results showed that the extent of policy influence was not as great as implied by the 1999 Review assessments. The change in explanatory power (adjusted R^2) when policy variables were included suggests that State policies have the tendency to enhance participation rates by about 8 per cent. We note, however, that the inclusion of a meaningful variable for VET in Schools may change this outcome.

---

15 See for instance the study by Waters, A., Greenwell, H. and Percival, R. (2002) *Education Participation Study,* National Centre for Social and Economic Modelling, University of Canberra.

16 Although there may be general consistency in signs, the coefficients for State variables in 'intercept and non-intercept models' generally differ. The omission of intercept affects the sizes of coefficients.

17 The alteration in the regression line or the association between the dependent and independent variables in most instances is reflected by a higher R^2 (cf. www.science.mcmaster.ca/geo/faculty/harris/geography/3z3dat.html).

THE COMPLETION RATE MODEL (LAMB)

Background

47. Following the November staff conference, Victoria sent the Commission a paper by Dr Stephen Lamb entitled "*Analysis of factors influencing post-compulsory enrolments using 2001 Census data*". Dr Lamb visited the Commission in early December 2002 to further explain his paper to staff and presented the paper to the Commission itself at the bilateral meeting in Victoria. In summary, the paper proposed a regression model using school completion rates for 19 and 20 year olds to analyse influences on the level of post-compulsory enrolments.

The approach proposed by Lamb

48. Lamb proposed a regression model to explain post-compulsory completion rates. He used the percentage of 19 to 20 year olds who had completed Year 12 as the dependent variable.

49. The independent variables consisted of:

(i) *Indigeneity* (represented by the percentage of Indigenous school students by SLA);

(ii) *Ethnicity* (a coefficient was derived to represent the situation where both parents were born in a non-English speaking country);

(iii) *School sector* (percentage of secondary school-age students in government, Catholic and independent schools in each SLA. This was a proxy measure because the Census did not record the sector of education in which 19 and 20 year-olds completed their secondary education);

(iv) *Location* (based on remoteness classification (ARIA) Lamb assigned 0=Major city; 1=Inner regional; 2=Outer regional; 3=Remote; 4=Very remote);

(v) *Socio-economic status* (measured using the SLA aggregate of the 1996 SEIFA for education and occupation);

(vi) *Part-time students* (State differences in the numbers of part-time senior school students were represented using the number of part-time secondary schools students by SLA); and

(vii) *State policy* (binary dummy variables were used to represent State-specific policy influences).

50. With this combination of variables, Lamb estimated a model that explained influences on post-compulsory enrolments to a significantly larger degree than the model which was the basis of the Commission's assessment in the 1999 Review. The model results showed that with the exception of part-time students, all the independent variables had a significant influence on the rate of completion of 19 and 20 year olds. The non-policy variables explained over 75 per cent of the variability in Year 12 completion. Introduction of State policy variables improved the explanatory power by 8 per cent. On the basis of this analysis, Lamb concluded that there was a strong case for the use of actual enrolments to obtain an accurate assessment of relative needs of States.

Re-specification of the completion rate model

51. We ran modified versions of the Lamb Model. Work involved in the re-specification of the model included testing the effects of:

(i) adjusting completion rates for mobility;

(ii) inclusion of specific non-policy variables in place of SEIFA;

(iii) using specific policy variables to replace or supplement the State dummies; and

(iv) inclusion of an intercept term.

52. ***Accounting for mobility.*** As acknowledged by Lamb, the mobility that is observed in the population cohort used in this study should be accounted for using information on previous place of residence.

53. We adjusted for mobility by deducting the number of 19 to 20 year olds who lived in a different SLA a year prior to the ABS 2001 Census from both the total population of this cohort as well as from the completion rate figures. Ideally, people who have moved in the year prior to the Census would have been re-assigned to their previous place of residence – but available data did not allow this.

Results and interpretation

54. The results of Lamb's original models (Models A and B) and the rerun of them (Models C to F) are in Table 4.

55. Models C and D performed better than the corresponding original models (compare with Models A and B). The improvement in the adjusted R^2 between Models A and C was 0.183 and that between Models B and D was .125. Models E and F had the following changes:

(i) inclusion of the intercept; and

(ii) the adjustment of these models for mobility.

The explanatory power of these models appears lower than that of the no-intercept models. As explained above, however, the two sets of results are not comparable.

56. As noted before, the adjustment of data for mobility was only partial. It involved deducting the number of 19 and 20 year-olds who lived in another SLA in the year prior to the Census year. The data we obtained from the 2001 Census was not detailed enough to enable re-allocation of those who lived in other SLAs one year ago to their original SLAs. We think that model performance would improve with this adjustment.

Table 4 RESULTS OF THE RERUN OF DR LAMB'S MODEL

	Model A	**Model B**	**Model C**	**Model D**	**Model E**	**Model F**
Non-policy						
Intercept					24.881	26.278
Remoteness	-2.55**	-2.50**	5.339**	4.764**	-1.970**	-4.891
Indigenity	-0.36**	-0.25**	-0.406**	-0.384**	-0.775**	-0.617**
Ethnicity	0.18**	0.26**	0.835**	0.711**	0.553**	0.560**
SEIFA Scale	0.10**	0.12**				
Catholic Schools	0.23**	0.19**	0.514**	0.362**		
Independent Schools	0.16**	0.03	0.492**	0.374**		
Non-government					0.500**	0.432**
Family Income			0.034**	0.031**	0.007**	0.017**
Full-time employment					-0.125*	-0.130
Policy						
Vic		3.59**		10.109**		2.552
Qld		11.66**		18.663**		11.695**
SA		-0.97		4.833**		-3.091
WA		-1.29		4.515**		-9.247**
Tas		2.20		2.383		-15.771**
NT		-5.65*		-3.275		-11.332**
ACT		1.88		23.612**		13.745 **
Proportion of part-time students		-0.89		-1.07**		0.699
Adjusted R^2	**0.751**	**0.833**	**0.934**	**0.958**	**0.316**	**0.385**

Notes: Model A = Original model presented by Dr Lamb;

Model B = Lamb's original model with policy variables;

Model C = Re-run of Model A replacing SEIFA with specific non-policy variables;

Model D = Re-run of Model B replacing SEIFA with specific non-policy variables;

Model E = Re-run of Model A replacing SEIFA and with adjustment for mobility (the corresponding model without intercept explained 78.0 per cent of the variability of 19 to 20 year-olds who completed post-compulsory education.); and

Model F = Re-run of Model B replacing SEIFA and with adjustment for mobility (the corresponding model without intercept explained 80.6 per cent of the variability of 19 to 20 year-olds who completed post-compulsory education).

** significant at 1% level of probability. * significant at 5% level of probability.

MODEL BASED ON NATIONAL AVERAGES

57. Staff used the 2001 Census data to calculate standardised enrolments by applying national average participation rates for the 15 to 20 year old population. Tables 5 and 6 show the national averages for the State populations disaggregated by age, location, Indigeneity and English speaking background.

58. Standardised enrolments were calculated separately for 15-17 year olds and for 18-20 year olds and the two results summed to derive total post-compulsory enrolments. This approach was adopted because of the marked differences in participation levels and in the pattern of participation between these two age groups. Standardised enrolments based on national averages were calculated by applying the averages in Tables 5 and 6 to the 15 to 17 and 18 to 20 year-old populations respectively in each of the five ARIA locations in each State. Table 7 shows the results compared with the actual enrolments as recorded in the Census and enrolments distributed on an equal per capita basis.

Table 5 NATIONAL AVERAGE PARTICIPATION RATES (PER HUNDRED STUDENTS) FOR THE 15 TO 17 YEAR-OLD POPULATION

ARIA Location	Indigenous students aged 15-17 years	Non –English speaking background students aged 15-17 years	All other students aged 15-17 years
Major City	55.93	83.01	75.85
Inner Regional	56.98	80.94	75.73
Outer Regional	54.85	77.28	72.46
Remote	50.76	61.39	67.72
Very Remote	31.51	30.83	55.02

Table 6 NATIONAL AVERAGE PARTICIPATION RATES (PER HUNDRED STUDENTS) FOR THE 18 TO 20 YEAR-OLD POPULATION

ARIA Location	Indigenous students aged 18-20 years	Non –English speaking background students aged 18-20 years	All other students aged 18-20 years
Major City	23.63	20.79	5.77
Inner Regional	19.86	33.01	7.04
Outer Regional	13.53	32.34	5.47
Remote	5.28	33.33	3.25
Very Remote	3.23	25.00	1.49

Table 7 STANDARDISED ENROLMENTS OF 15 TO 20 YEAR-OLDS BASED ON AVERAGE PARTICIPATION RATES

	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	Aust
Actual enrolments	219387	174706	116644	57141	51652	13976	12987	5182	651675
Standardised enrolments (National average model)	217685	160854	124088	66178	49758	16300	11521	5291	651675
Enrolments based on an equal per capita distribution	214446	159353	126175	66726	49627	16053	12210	7085	651675

CALCULATION OF STANDARDISED ENROLMENTS

59. The following sections describe the use of the regression models to calculate standardised enrolments.

Standardised participation

60. The results of Model 3 (Table 3) were used to estimate total standardised post-compulsory enrolments. The constant and the non-policy variables were used in the calculation. The policy variables and the variation in participation not explained by the model have been ignored. We only wish to standardise enrolments for identifiable non-policy influences. This could be interpreted as implying that all the unexplained variation is related to policy differences. This probably results in a conservative assessment of disabilities but errs on the side of policy neutrality. Attachment C provides more details of how results produced by the model were used to derive standardised enrolments.

61. Table 8 compares actual enrolments recorded in the 2001 Census with the standardised results from the participation rate model and with the equal per capita distribution of enrolments.

Table 8 STANDARDISED POST-COMPULSORY ENROLMENTS

	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	Aust
Actual enrolments	219387	174706	116644	57141	51652	13976	12987	5182	651675
Standardised enrolments (Participation rate model)	216501	164797	122590	64349	50139	15383	12630	5286	651675
Enrolments based on an equal per capita distribution	214446	159353	126175	66726	49627	16053	12210	7085	651675

Standardised completion

62. The method of calculating standardised completion for States is described in Attachment C. The calculation is based on the coefficients in Model F in Table 4. Table 9

compares actual completions recorded by the 2001 Census with the standardised results from the completion rate model and with the equal per capita distribution of completions.

Table 9 STANDARDISED COMPLETIONS FOR 19-20 YEAR OLDS BASED ON THE COMPLETION RATE MODEL

	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	Aust
Actual completions	107447	84547	68773	32092	23316	6346	8507	2007	333035
Standardised enrolments (Completion rate model)	114143	90972	57320	31320	24072	6200	6580	2427	333035
Completions based on an equal per capita distribution	108771	82617	64394	33732	24966	7755	6976	3824	333035

Comparison of results

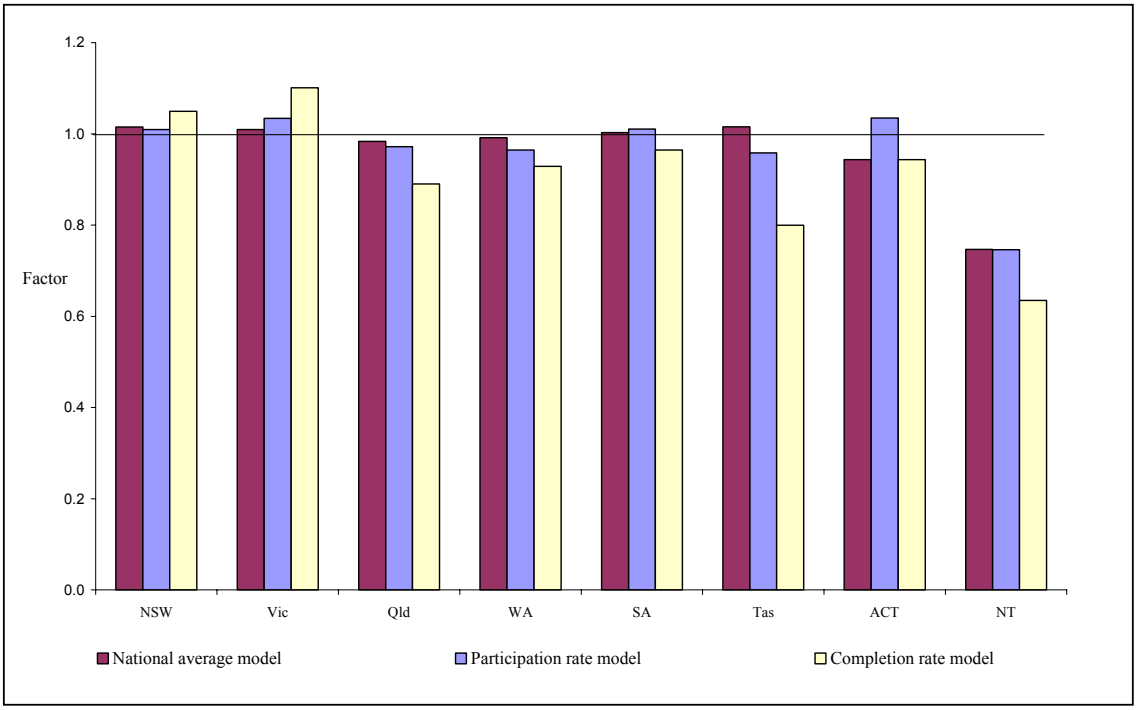
63. Direct comparison of the two estimations of standardised enrolments based on participation and the completion rates was not possible. However it was possible to compare the factors implied by the standardised results. A comparison of these factors is shown in Table 10 and in Figure 2. Table 10 also includes the results obtained from applying national averages (see Table 6 above).

Table 10 COMPARISON OF ADJUSTMENT FACTORS

Standardised enrolments	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	Aust
Participation rate models	1.00958	1.03416	0.97159	0.96437	1.01034	0.95828	1.03435	0.74605	1.00000
National average model	1.01510	1.00942	0.98346	0.99178	1.00265	1.01543	0.94351	0.74680	1.00000
Completion rate model	1.04939	1.10113	0.89015	0.92849	0.96419	0.79950	0.94323	0.63472	1.00000
Actual enrolments	1.02304	1.09634	0.92447	0.85635	1.04081	0.87064	1.06360	0.73136	1.00000

64. Figures in Table 10 enable comparison to be made of the relative size and direction of adjustment due to each of the models discussed in the paper. Compared to the other models, the factors derived from the participation rate model are more comparable with the factors implied by actual enrolments. The factors are in the same directions for all States — but in virtually all cases the factors derived for the participation rate model are closer to unity. Nevertheless when compared with the National average model, the participation rate model showed marginal inconsistency in the direction of change of factors for Tasmania and the Australian Capital Territory.

Figure 2 COMPARISON OF ADJUSTMENT FACTORS FOR NATIONAL AVERAGE, PARTICIPATION RATE AND COMPLETION RATE MODELS



65. The completion rate model produced markedly different results for most States. The results for Tasmania, the Australian Capital Territory and the Northern Territory are the most divergent. These results may suggest that further attention needs to be paid to the variables considered in this model — in particular the effects of mobility of the 19-20 year old population within the completion rate model.

IMPLICATIONS FOR THE 2004 REVIEW

66. Within the constraints of the available data, staff consider that the participation rate analysis presented here provides a reasonably robust and policy neutral basis for assessing the affect of non-policy influences on post-compulsory enrolments. We think that this analysis provides a basis that is superior to the alternatives because it is based on the wide-ranging data describing socio-demographic status available in the ABS Census at the SLA level.

67. We think that the results produced by the participation rate analysis are more suitable for our purposes that those derived from the completion rate model because it is difficult to remove, from the data used in that model, the affects of mobility of the post

school age population. Based on the outcome from the present analysis, we expect that if this difficulty could be overcome, the results from the completion rate analysis would be broadly supportive of those obtained from the participation rate analysis — the ranking of the States would be similar.

68. We also think that the results of the participation rate analysis are preferable to those obtained from the application of national average participation rates. They are both methods of approximation. The regression based model has the capacity to take more precise account of the variability within the underlying data and a much wider range of influences causing such variability than is possible in the national average model.

69. Staff propose that the results of the participation rate analysis (as adjusted to include a VET in schools variable) be used in the assessment of socio-demographic composition for secondary education for the 2004 Review. Table 11 shows the results of using the adjustment factor derived from the current participation rate model (see Table 8 above) as input into the socio-demographic factor for the government secondary education category. Table 12 shows the corresponding figures for the non-government secondary education category. The post-compulsory adjustment factors have been included in the calculation of the socio-demographic composition factors using the same method as in the 1999 Review. This method is outlined in Attachment C.

Table 11 SOCIO-DEMOGRAPHIC FACTORS FOR GOVERNMENT SECONDARY EDUCATION, 2001-02

Standardised enrolments	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	Aust
2003 Update	0.96024	0.89076	1.11545	1.17156	0.98273	1.07912	0.95852	1.21887	1.0000
Based on participation rate model	0.96398	0.90950	1.10786	1.13869	0.98884	1.04734	0.98147	1.08596	1.0000
Based on actual enrolments	0.97122	0.93923	1.08407	1.08225	1.00435	0.99695	0.99788	1.07659	1.0000

Table 12 SOCIO-DEMOGRAPHIC FACTORS FOR NON-GOVERNMENT SECONDARY EDUCATION, 2001-02

Standardised enrolments	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	Aust
2003 Update	0.92793	1.00161	1.08692	1.16818	0.94127	0.77457	1.21686	0.77490	1.0000
Based on participation rate model	0.93013	1.02164	1.07743	1.13305	0.94588	0.75152	1.24010	0.70911	1.0000
Based on actual enrolments	0.93593	1.05449	1.05169	1.07442	0.96012	0.71592	1.25637	0.70366	1.0000

70. Table 13 shows the redistribution of grants amounting to $64.9 million as a result of the changes in the standardised student numbers in government secondary and

non-government secondary categories. The amounts redistributed were estimated using the State Grant Simulator.

Table 13 REDISTRIBUTION OF GRANTS, 2003 UPDATE ($M)

Standardised enrolments	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	Aust
2003 Update	12.1	46.0	-19.2	-27.0	3.5	-6.3	3.3	-12.2	64.9

Feedback from States

71. We seek comments on all aspects of this work. Comments are sought on:

(i) the suitability of the presented models for measuring the impact of non-policy effects on post compulsory enrolments; and

(ii) the integration of the results of the modelling into the assessments.

ATTACHMENT A

COMPARISON OF DATA SOURCES

1. Table A1 provides a comparison between the total numbers of students shown in *Schools Australia* and the numbers derived from the 2001 Census.

Table A1 COMPARISON OF CENSUS DATA WITH *SCHOOLS AUSTRALIA* DATA

Age	15	16	17	18	19	20	Total
Government Schools							
Schools Australia publication	161 358	137 153	102 835	21 162	3 088	3 513	429 136
ABS 2001 Census	147 359	128 735	100 232	26 412	4 047	1 194	407 979
Non-government Schools							
Schools Australia publication	87 708	79 722	64 521	12 296	1 204	414	245 865
ABS 2001 Census	84222	78004	64099	15061	1776	534	243 696
Total School Sectors							
Schools Australia publication	249 066	216 875	167 356	33 458	4 292	3 927	674 974
ABS 2001 Census	231 581	206 739	164 331	41 473	5 823	1 728	651 675

Notes: The cross-tabulated ABS 2001 Census data summarised in the table were used in the regression model. In this summary, data for only 1350 statistical local areas were used as in the regression analysis.

Data reported in the ABS *Schools Australia* publication of 20th June 2001 (4221.0) was used for the comparison.

ATTACHMENT B

INDEPENDENT VARIABLES IN THE MODEL

Introduction

1. An individual's capacity and willingness to extend learning are influenced by socio-demographic status, location and government policies. Independent variables are the measures that permit the analysis of these influences within a modelling framework. This section provides detailed specifications of such measures — especially:

(i) the literature based justification for the variables used;

(ii) the form in which the variables will be included; and

(iii) the likely interactions between the variables.

2. When defining variables, Commission staff took account of the need to avoid multi-collinearity between independent variables. Therefore, through preliminary analysis, any formulations of independent variables that were highly dependent on one another to the extent of undermining the significance of individual coefficients were avoided[18]. Formulations of variables that directly competed with the dependent variable[19] (employment/unemployment, vocational education and training etc) were also avoided.

3. Table B1 shows the full list of generic independent variables that formed the basis for modelling.

Table B1 INDEPENDENT VARIABLES USED IN THE ANALYSIS

Non-policy variables	Policy variables
Family income	
Ethnicity	Average age of students in Year 10
Indigeneity	Proportion of part-time students
Remoteness	School type
Employment opportunities	Participation in VET
School sector	

Note: A further policy variable for VET in schools will be included in a revised model when data are available and after State feedback has been received on the work outlined in this paper.

[18] For this purpose, widely available statistical tests — including analysis of variance were employed.

[19] These are variables that could be used as an alternative dependent variable.

Non-policy variables

4. Many of the non-policy variables considered in the analysis were socio-economic variables and served as measures of an individual's relative position in the community with regard to income, occupational and educational backgrounds. The other variables considered were largely of a demographic nature.

5. ***Family income, educational and occupational status***. A large body of literature and the Commission's research indicated that socio-economic status of parents — specially their educational and occupational background — are significant drivers of post-compulsory schooling[20]. For example, regression analysis undertaken by the National Centre for Social and Economic Modelling (NATSEM) [21] at the University of Canberra showed that the high socio-economic status of parents had a statistically significant and positive relationship with high educational participation for students aged 15-17. Family income is associated with the ability to fund post-compulsory schooling — the higher the family income, the higher the ability of a family to support their children to pursue post-compulsory education (Lewis and Koshy, 1999)[22]. Research carried out by Lamb, Dwyer and Wyn (2000)[23] and Ainley and Sheret (1992)[24] supported this. In particular, these studies have shown that early school leavers were much more likely to come from lower socio-economic backgrounds. Some of the State submissions have also indicated the importance of including this variable.

6. For purposes of modelling, the inclusion of all aspects that reflect socio-economic status could result in multi-collinearity. As family income is generally regarded to be closely correlated with educational and occupational background, it was chosen as a proxy for all the socio-economic variables relevant to our model. A number of different forms of family income have been used in regression models. DEETYA (2000:p14)[25] noted the use of weekly earnings as 'a proxy for family income'. In this model, family income was represented as the average income for families in each SLA using the mean of the weekly incomes of families supporting post-compulsory students aged 15 to 17 years (see the following formula). Data for the relevant families were obtained from cross-tabulations of the ABS 2001 Census.

---

20 See for instance, Marks, G. N., Fleming, N., Long, M. and McMillan, J. (2000) *Patterns of participation in Year 12 and higher education in Australia: Trends and issues, p 43,* Research Report Number 17, Australian Council for Education Research, DEETYA, Canberra. Their literature indicates that 'the effect of parental occupation on educational participation has declined' over the recent years.

21 Waters, A., Greenwell, H. and Percival, R. (2002) *Education Participation Study*: National Centre for Social and Economic Modelling, University of Canberra.

22 Lewis, P.E.T. and Koshy, P. (1999) Youth employment, unemployment and school participation, *Australian Journal of Education,* 43(1), pp 42-57.

23 Lamb, S., Dwyer, P. & Wynn, J. (2000) *Non-completion of school in Australia: the changing patterns of participation and outcomes* (LSAY Research Report No. 16*)*. ACER: Melbourne.

24 Ainley, J. and Sheret, M. (1992) *Progress Through High School: A Study of Senior Secondary Schooling in New South Wales,* Research Monograph No. 43 Melbourne: Australian Council for Educational Research.

25 DEETYA, (2000) *Participation in post-compulsory schooling,* IAED Occasional Paper Series 3/2000, Research and Evaluation Branch, Department of Education, Training and Youth Affairs, Canberra.

$$\text{Average Family income} = \frac{\Sigma[\text{mean family income for the income bracket} * \text{number of relevant families in that bracket}]}{\text{Total number of families in the SLA}}$$

7. ***Ethnicity.*** In the Australian context, the issue of education and ethnicity is extremely important because a quarter of the population was born in other countries. Recent studies have highlighted the important links between post-compulsory education and the ethnicity of students (cf. Tees 2002)[26]. Lamb (1998), Lamb, Dwyer and Wyn (2000) and Williams (1987) have shown the presence of consistent differences in school completion by children of parents who were born in non-English speaking countries compared to the general population.

8. As used in empirical studies both here and overseas, this model used 'language spoken at home' as a proxy for low fluency in English. Based on preliminary analysis of the relationship between participation rate and the ethnicity rate, the following variable was constructed by taking the percentage of 15 to 17 year-olds who speak a language other than English at home (equivalent of NESB) participating in post-compulsory education in a SLA relative to the total 15 to 17 year-olds in that SLA. This specification of the variable was preferred to one based on the total number of NESB persons or families in an SLA. We expected that there might be marked differences between the proportion of NESB persons in an SLA as a whole and the proportion of NESB persons in the post compulsory age group.

$$\text{NESB (15-17) rate} = \frac{\text{Total 15 to 17 year-old NESB equivalents in post-compulsory education}}{\text{Total 15 to 17 year-old NESB equivalents in the SLA}} * 100$$

9. The variable used in the model for the 18 to 20 year-olds was formulated using a similar procedure.

10. ***Indigeniety.*** Three of the four States with the highest proportion of Indigenous people in their populations have the lowest participation rates in post-compulsory schooling (CGC, 2001:p13)[27]. Lamb (2002) noted that the rate of enrolment of Indigenous students in Year 12 is 30 to 40 per cent lower than that of non-Indigenous students.

11. A proxy variable was constructed as shown below.

$$\text{Indigenity} = \frac{\text{Total Indigenous persons in a SLA}}{\text{Total persons in a SLA}} * 100$$

12. ***Remoteness.*** Based on widely available empirical literature as well as our own research, we concluded that students outside the metropolitan areas have reduced

---

[26] Prof. Richard Tees' work as reported in the *University of Melbourne – News Release* of 3rd September 2002.

[27] CGC (2001) *Issues in the Schools Education Assessments for the 2004 Review*, Discussion Paper CGC 2001/17, Commonwealth Grants Commission, Canberra.

access to post-compulsory education[28]. Highly urbanised States are likely to have higher participation rates. Similarly, less urbanised or remote locations can be expected to have lower participation rates.

13. On the assumption that location and population density are closely correlated, the influence of remoteness on post-compulsory enrolments was depicted in the model using the ABS standard geographical classification based on the Accessibility/Remoteness Index of Australia[29] (ARIA).

14. ABS classified SLAs into 5 groups using these ARIA scores. Our preliminary analysis of the correlations between post-compulsory participation and remoteness showed that the sizes of the coefficients for different levels of remoteness were similar. Therefore in subsequent modeling, only two variations of remoteness were used. The SLAs grouped as 'very remote' according to ARIA were assigned a value of 1. The remaining SLAs were assigned zero. Therefore this variable was meant to test whether remoteness had a negative influence on participation.

$$\text{Remoteness} = \begin{cases} 1 \text{ for all SLAs that are very remote} \\ 0 \text{ otherwise} \end{cases}$$

15. The most up-to-date data for ARIA were not available in a form that was suitable for the current analysis. Therefore the required data for this variable were obtained from publications of the 1999 update of ARIA[30].

16. In this analysis, location has been treated as a non-policy variable. If it is recognised that geographical isolation compounds the influence of socio-economic disadvantages, then for modelling purposes location can be treated as a non-policy variable. A case could be made, however, that to the extent that differences in participation in non-metropolitan areas are due to different levels of access to schools, location could equally well be treated as a policy variable.

17. ***School sector.*** Available evidence suggested the presence of a strong positive correlation between socio-economic status and the choice of schools. Mukherjee (1999)[31] noted that a greater proportion of students in government schools were from the lower half of the socio-economic spectrum. This he attributed to the fact that government schools are free or their fees were minimal compared to schools in other sectors. A similar trend was observed in the enrolments in Catholic schools where the fees were relatively low (Mukherjee, 1999). Other non-government schools appeared to attract

---

28 Although the extent to which this is the case varies between States and may be due to policy differences. See also the work by Lamb, Dwyer and Wyn, 2000; Tees, 1999.

29 ARIA is an aggregate measure of the road distance to service centers in each Statistical Local Area (SLA) and ranges from zero for high accessibility to 12 depicting extreme remoteness. There are two versions of ARIA. In our analysis we used the ABS version (ARIA plus).

30 The most up-to-date ARIA classification is available at CD-level. It has not been used in this analysis due to the amount of work involved in translating it to SLA-level.

31 Mukherjee, D. (1999) *Socio-economic Status and School System Enrolments*, Australian Centre for Equity through Education, http://members.ozemail.com.au/~devm/SES.HTML.

students from high socio-economic backgrounds. A recent study by Lamb, Hogan & Johnson (2001)[32] confirmed these observations. Therefore capturing sector influence when modelling post-compulsory schooling is warranted.

18. Cross-tabulated ABS 2001 Census data on school sector for the relevant ages were used in constructing the school sector variable. In the past, researchers have used a range of variables to represent school-sector influence. They formulated these variables by taking the ratio of post-compulsory school enrolments in each school sector. To avoid likely incidence of multi-collinearity due to this approach[33], we used a single variable to represent the school-sector influence. For the participation rate model, the variable was constructed by taking the total enrollments of 15 to 17 year-olds in non-government schools in a SLA as a ratio of the total enrolments in the corresponding age group in all school sectors as shown below.

$$\text{Non-government} = \frac{\text{Total enrolments in non-government secondary schools in a SLA}}{\text{Total enrolments in all schools in the SLA}} * 100$$

19. ***Employment opportunities.*** The availability of full-time employment opportunities has a significant bearing on post-compulsory schooling (Roussel and Murphy, 2000:p22). The work by NATSEM for the ACT Treasury suggested that for Australia as a whole, educational participation was significantly and inversely correlated with the proportion of 15–19 year olds who were employed full-time and positively correlated with youth unemployment rate. It is our view that the inclusion of both these variables simultaneously in a regression could lead to multi-collinearity and other problems. In our model we captured the effects of the availability of opportunities for full-time employment on post-compulsory participation as follows.

$$\text{Full-employment} = \frac{\text{Total number of employed persons between 15 and 19 years of age in a SLA}}{\text{Total number of 15-19 persons in the SLA}} * 100$$

20. To formulate this variable, we used employment data for 15 to 19 year-olds in a SLA because the required data for the relevant age group were not available[34].

Policy variables

21. Some States have argued that differences in State policies have only a marginal impact on participation rates. They have implicitly attributed any differences that arise to socio-economic and other similar influences. Some researchers like Williams,

---

[32] Lamb. S. Hogan, D. and Johnson, T. (2001) The stratification of learning opportunities and achievement in Tasmanian secondary schools, *Australian Journal of Education,* 45(2), 153-167.

[33] Formulation of separate variables require the use of a common denominator — consequently introducing multi-collinearity.

[34] Data required for the construction of this variable were obtained from CDATA96 because comparable data are not yet available from the 2001 Census.

Long, Carpenter and Hayden (1993)[35] showed that such State differences played only a minor role in schooling decisions.

22. In a number of past research studies, State differences other than those due to socio-economic and other similar influences were captured using dummy variables. In the modelling we have undertaken, specific variables were included in the model to capture the effects of State policies which might be influencing participation in post-compulsory education. Capturing these effects in simply calculated and pertinent variables proved to be difficult. In initial runs of the model, the explanatory power of these variables was not strong. In the final version of the model, State dummies were included to represent more comprehensively the influences of State policy effects.

23. ***School type.*** To test the influence of the type of institution providing post-compulsory education on participation, we examined data on school types provided by the States in the Commission's Education Special Data Collection. We decided to use the percentage of students attending secondary colleges in a statistical sub division (SSD) as a proxy for the school-type variable. Commission research suggest that secondary colleges are an important driver of participation. SSDs were used as the basis for this variable because the spread of the schools was not uniform at the SLA level.

24. Different school types available for post-compulsory education are shown in Table B2 and the algebraic formulation of the variable is shown below.

Table B2 SCHOOL-TYPES AVAILABLE FOR POST-COMPULSORY EDUCATION

Variable	Definition
School-type 1	Schools designed to provide all years of secondary schooling
School-type 2	Schools (other than Secondary Colleges) which are not designed to provide all years of secondary education
School-type 3	Secondary College
School-type 4	Technical High School
School-type 5	Other types of schools such as mixed primary/secondary schools, bilingual Aboriginal secondary schools, Aboriginal homeland centres, other Aboriginal schools etc

$$\text{School type} = \frac{\text{Total number of secondary students attending type 3 schools in a SSD}}{\text{Total number of secondary students in the SSD}} * 100$$

25. ***Average age entering post-compulsory schooling.*** The Commission research showed that the rate of participation in post-compulsory schooling was associated with the age of commencement of school. Consequently, the States that have a lower age of

[35] Williams, T., Long, M., Carpenter, P. & Hayden, M. (1993) *Year 12 in the 1980s*. Canberra: Australian Government Publishing Service.

commencement tend to have a lower participation rate[36] and lower average age at Year 10. Two specifications of a variable to measure this effect were tested — commencement ages (5 and 6) and the average age of Year 10 students in each State. The final version of the model uses the second variable.

26. Table B3 shows the average age of students in Year 10 in each State. It can be inferred from the data in the table, that States like Queensland, South Australia and Western Australia tend to have relatively lower rates of participation. In these States, the commencement age of secondary schooling is lower than the average age of commencement of post-compulsory schooling of about 15.6 years. In our model, this influence was captured through a binary dummy variable. The States with a lower participation rate relative to the average were assigned a value of zero and the others a value of 1.

$$\text{School start} = \begin{cases} 1 \text{ for all states where the start age was above the Australian average of 15.6} \\ 0 \text{ otherwise} \end{cases}$$

Table B3 AVERAGE AGE OF STUDENTS BEGINNING POST-COMPULSORY EDUCATION

State	Average age
NSW	15.70
Vic	15.76
Qld	15.24
WA	15.54
SA	15.12
Tas	15.65
ACT	15.85
NT	15.62

Source: ABS Catalogue 4221.0 – Schools, Australia.

27. ***Part-time students.*** State policies and organisational arrangements affect the percentage of part-time students (P) in each jurisdiction (ABS 2002)[37]. A variable for part-time students was included to test the extent to which differences between States in the availability of this option affected participation rates.

28. The variable was calculated using cross-tabulated data obtained from the 2001 Census. In a similar manner to Lamb's study, the influence of 15 to 17 and 18 to 20 year-old part-time students were modelled by taking the percentage of part-time secondary school students in these groups in an SLA.

29. Denoting the influence of part-time students by P, the variable used to model participation by 15 to 17 year-olds was defined as:

---

[36] Western Australia in its '*Rejoinder Submission'* stressed this point.

[37] For details see, Catalogue No. 4221.0 — 'Schools, Australia', Australian Bureau of Statistics, 2001.

$$\text{Part-time} = \frac{\text{Part-time secondary school students of a particular age group in a SLA}}{\text{Total secondary school students in that SLA}} * 100$$

30. The variable used to model participation by 18 to 20 year-olds was defined in a similar way.

31. ***Participation in VET.*** Vocational education and training (VET) has two important influences on post-compulsory participation. VET in schools has an augmenting effect on participation whereas VET outside school system has a negative effect. Ideally, the influence of both these aspects should be explored.

32. *VET in schools.* The ACT in its submission said that VET in schools is an important issue in assessing post-compulsory school participation. It noted that VET in schools has boosted post-compulsory school participation in all jurisdictions and was an attractive alternative option to be considered by those who have completed the compulsory years of schooling. Table B4 shows a comparison of the rates of participation in school industry programs and post-compulsory participation. This is at best a rough indication because the definition of industry programs in the source used here is probably not comparable with the definition of VET in school courses.

Table B4 PARTICIPATION RATES IN SCHOOL INDUSTRY PROGRAMS AND IN POST-COMPULSORY EDUCATION

	In school industry programs	Participation rate of 15 to 17 year-olds
NSW	23.7	75.1
Vic	10.4	80.7
Qld	21.7	71.6
WA	18.0	53.9
SA	22.7	48.3
Tas	12.2	61.7
ACT	73.2	76.0
NT	18.1	51.2

Source: VET in Schools data is in Malley, J., Ainley, J. and Robinson, L. (2001) *Witnessing Evolution – Student participation in programs*, Australian Council for Education Research (ACER) (see, www.ecef.com.au/web/km/kmgateway.nsf/ecef/Student_participatio4YL3HC?OpenDocument).

Participation in post-compulsory education from the ABS 2001 Census.

33. It has been observed that students wishing to undertake a vocational education now have two options: VET provided in secondary schools and vocational courses in TAFE institutes. A significant trend towards opting for school-based VET could be expected to show a positive influence on post-compulsory schooling.

34. We explored a number of key sources[38] to obtain VET in schools data. We understand that VET in schools data at the State level will become available from MCEETYA in July. Ideally, we need data at the school level or aggregated by area (SLA or postcode) to enhance the model. Staff are writing to State Treasuries to seek access to school level data on levels of participation in VET in school courses.

35. Available information suggests that States have adopted varying definitions of school level VET.

36. *VET outside the school system.* To test whether outside VET had a negative influence on participation in post-compulsory schooling, a variable was included by taking the VET contact hours per capita for all participants in VET in each SLA as shown below.

$$\text{Outside VET} = \frac{\text{Total contact hours associated with outside VET in a SLA}}{\text{Total persons in that SLA}} * 100$$

37. This variable was common to participation models for both age groups.

State effects

38. State effects arise due to different impacts of the policies identified above and of other unspecified State policies. Interactions between State policy influences and non-policy influences could also lead to significant State effects. It is difficult to identify and quantify all of these effects. Therefore in our model, the State influences were represented using a set of binary dummy variables. An example of how State effects were introduced into the model for Victoria (Vic) using a binary dummy variable is shown below:

$$\text{Vic} = \begin{cases} 1 \text{ for all SLAs in Victoria} \\ 0 \text{ otherwise} \end{cases}$$

39. To avoid the 'dummy variable trap', all States but one were included in the model. The variable to be dropped was determined by the statistical package that was used in the regression analysis.

Intercept of the model

40. As noted before, it was necessary to capture the influence of State effects on participation and completion rates. Adhering to econometric guidelines, these effects were represented in the participation and completion models by using dummy variables for all

---

[38] The sources explored include the Ministerial Council on Education, Employment, Training and Youth Affairs (MCEETYA); the National Centre for Vocational Education Research (NCVR); and the Australian National Training Authority (ANTA).

States but one. The omission of one of a set of dummy variables as in this instance affects the intercept term of the models. When using these models to estimate standardised enrolments, it was necessary to adjust the value of the intercept term by either adding or subtracting an amount equivalent to the quantified effect of the omitted dummy (see Suits, 1984[39]). We used Suits' approach when adjusting the intercepts of Model 3 and Model F. In both cases, the dummy variable for NSW was omitted when the models were run.

41. Derivation of the adjustment required for Model 3 is illustrated below.

$$\text{Adjusted intercept of Model 3} = \text{Original intercept of Model 3} \pm \frac{\text{Sum of the coefficients for the 7 State dummies}}{8}$$

where, 8 represents the total number of States considered in each model.

42. Table B5 shows the derivation of the adjusted intercepts for the participation rate and completion rate models.

Table B5 ADJUSTED INTERCEPTS

	Original Intercept	Adjustment	Adjusted Intercept
Participation Rate Model 3	40.1550	+ 1.5740	41.7290
Completion Rate Model F	26.2780	- 1.4240	24.8540

[39] See for instance Suits, D. B. (1984). Dummy variables: Mechanics v. interpretation, *Review of Economics and Statistics*, 68(1), 177-180.

ATTACHMENT C

MODEL-BASED ESTIMATION OF STANDARDISED ENROLMENTS

1. This attachment shows how the results of the Model 3 were used to calculate standardised enrolments for each State. It also describes how these results were used to derive illustrative socio-demographic composition factors for Government and Non-Government Secondary Education.

2. The following box illustrates the application of Model 3 to calculate the standardised participation rate for each State[40].

$$SP1517_{SLA} = 41.729 - 7.139*\text{Remoteness} - 0.370*\text{Indigenity} + 0.021*\text{Ethnicity} + 0.072*\text{Non-government proportion (15-17 year-olds)} + 0.049*\text{Family Income (15-17 year-olds)} - 0.310*\text{Full-time Employment}$$

3. When using the above formula, the steps set out below were followed to calculate the State specific standardised post-compulsory enrolments. The steps are illustrated using the participation rate model for 15 to 17 year-olds:

(i) for each SLA, the actual value of each independent non-policy variable was adjusted by multiplying it by the coefficient produced by the model for that variable;

(ii) these adjusted values were summed together with the adjusted intercept value produced by the model to derive standardised participation rates for each SLA — ($SP1517_{SLA}$);

(iii) standardised enrolments for each SLA were calculated by multiplying the total population of 15 to 17 year-olds in that SLA by the standardised participation rate for the SLA derived at step (ii) — ($TP*SP1517_{SLA}$); and

(iv) the results for the *n* SLAs in State *j* were summed to obtain the State total S_j shown in the formula below.

$$S_j = \sum_{i=1}^{n} (TP*SP1517_{SLA})$$

where n = the number of SLAs in a State;

TP = total 15-17 year olds in the SLA; and

j = the State identification.

---

[40] The coefficients of policy variables are not included as we wish to exclude their influence on participation. However, the sizes of the coefficients for non-policy variables reflect the influence of the State policy variables.

4. A similar procedure was adopted to calculate standardised enrolments based on the completion rate model (Model F in Table 4).

5. Staff explored the following two approaches to estimate total standardised post-compulsory enrolments for inclusion in the socio-demographic composition factors for the Government and Non-Government Secondary Education categories.

(i) ***Method A***. Calculate standardised enrolments of 15 to 17 year-olds using Model 3 and apply the coefficients of this model to extrapolate the results to 18 to 20 year-olds. Aggregate the results for 15 to 17 year-olds and the extrapolated results for 18 to 20 year-olds.

(ii) ***Method B.*** Calculate standardised enrolments of 15 to 17 year-olds using Model 3. Run a separate regression for 18 year olds using the same set of parameters as used for Model 3 and use the results to calculate standardised enrolments of 18 year-olds. Apply the coefficients of the model for 18 year-olds to extrapolate the results to 19 to 20 year-olds. Aggregate the results for 15 to 17 year-olds, the results for 18 year-olds and the extrapolated results for 18 to 20 year-olds.

6. Table C1 shows the standardised post-compulsory enrolments derived under the two methods.

Table C1 STANDARDISED POST-COMPULSORY ENROLMENTS

	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	Aust[(a)]
Standardised enrolments Method A	216501	164797	122590	64349	50139	15383	12630	5286	651675
Standardised enrolments Method B	217478	164776	121734	63938	50166	15811	12340	5432	651675

(a) Total post-compulsory enrolments derived from the 2001 Census

7. Table C2 shows the adjustment factors derived using the two methods and also the adjustment factor used in the 1999 Review.

Table C2 COMPARISON OF THE ADJUSTMENT FACTORS UNDER THE TWO PROPOSED METHODS

	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	Australia
Method A	1.00958	1.03416	0.97159	0.96437	1.01034	0.95828	1.03435	0.74605	1
Method B	1.01413	1.03403	0.96481	0.95821	1.01087	0.98496	1.01063	0.76659	1
1999 Review	1.00070	1.00799	0.98642	1.01560	0.98758	0.96954	1.07584	0.92368	1

8. Table C3 shows the results of applying the adjustment factors set out in Table C2 to the total post-compulsory enrolments derived for 2001-02 from the ABS publication *Schools Australia*[41].

Table C3 COMPARISON OF ESTIMATED ENROLMENTS OF PERSONS AGED 15 AND OVER USING THE PROPOSED METHODS

	NSW	Vic	Qld	WA	SA	Tas	ACT	NT	Australia
Method A	228725	173208	129444	68399	52253	16191	13155	5582	686958
Method B	229756	173186	128540	67962	52280	16642	12856	5736	686958
1999 Review	227282	166711	131572	72908	51672	17133	12633	7047	686958

9. Table C4 shows the percentage change in enrolments between Methods A and B and the 2003 Update.

Table C4. PERCENTAGE CHANGE IN THE ADJUSTMENT FACTOR UNDER THE TWO METHODS

Type of change	NSW	Vic	Qld	WA	SA	Tas	ACT	NT
Between Method A & U2003	0.89	2.60	-1.50	-5.04	2.30	-1.16	-3.86	-19.23
Between Method B & U2003	1.34	2.58	-2.19	-5.65	2.36	1.59	-6.06	-17.01

10. Staff chose Method A as the basis of the adjustment factor included in the socio-demographic composition factor for the illustrative calculations shown in this paper. This was largely because it was thought that the mobility of 19 and 20 year-olds may be affecting the results that were obtained from Method B.

11. Using the adjustment factor derived from Method A, standardised post-compulsory enrolments were included in the assessment of socio-demographic composition in the following way:

(i) The adjustment factor for each State was applied to the standard number of enrolments (using ABS schools data) to derive the standardised number of students — both government and non-government, aged 15 years and over for each State.

(ii) Each State's notional number of standardised students aged 15 years and over was added to its actual secondary students aged 14 years and under (excluding year 7) and its notional Year 7 students, to derive the total standardised number of secondary students.

---

41 ABS *Schools Australia* (4221.0) publication of 20 June 2001.

(iii) Each State's resulting total standardised number of secondary students was apportioned between government and non-government sectors using its actual proportions in those sectors[42].

(iv) Separately for Government and Non-Government Secondary Education, socio-demographic factors were obtained by calculating secondary school students per capita for each State and Australia and by dividing each State's per capita value by the Australian value.

12. The method for deriving the socio-demographic composition factors described here is essentially that used in the 1999 Review.

---

[42] The Year 7 proportions were adjusted to take account of the influences resulting from State differences in sector placement (that is, primary or secondary).

ATTACHMENT D

COMMENTS IN REJOINDER SUBMISSIONS

1. New South Wales did not provide comment on these issues.

2. Victoria said that it supported the Commission's efforts to develop a new assessment that was policy neutral in the calculation of post-compulsory enrolments. It pointed out that the use of a 'proxy' to measure a range of socio-economic factors as the key weakness of the existing method. It said that it was encouraged by the Commission's recognition of its work in support of a more comprehensive treatment of the key non-policy determinants of post-compulsory enrolments.

3. Queensland argued that participation rates should be used as a measure of post-compulsory enrolments. This was because participation rates were more comprehensive measures and the fact that the use of retention rates would require fundamental changes in the way school enrolments were included in the assessment. Quoting Western Australia, it said that if retention rates were used, the Commission would 'have to use grade based groups' and be confident that the coverage of sample retention data used by Victoria were comprehensive enough to enable the same degree of reliability in the outcome as potentially available from participation rate data. It said that the use of participation rates on the other hand fitted 'better with other Census data used in the assessment' of post-compulsory enrolments.

4. Western Australia noted that the major advantage of the Commission's existing model was its policy neutrality – it was free of any impacts due to policy changes that affect post-compulsory enrolments. It supported this statement with evidence such as the significant influence that policy measures like commencement age had on its average post-compulsory participation rate. Western Australia said that that any changes to the model should not be distorted by inadequate model specifications and the way important variables like urbanisation and socio-economic status were defined.

5. South Australia said that the 'starting point of any post-compulsory enrolments should be Australian average participation rates'. It was concerned about the retention rate model proposed by Dr Lamb and included in Victoria's main submission — because it used full-time retention rates and did not account for part-time students when defining the dependent variable. This ignored an issue that was significantly important to South Australia. The fact that it implied that State policy influences would be captured by the State-specific (dummy) variables was another aspect of concern to South Australia. It pointed out the model's failure to recognise labour market conditions adequately, despite strong supporting research evidence about the association between employment opportunities and lower participation. It said that if the Commission's existing model for measurement of post-compulsory enrolments were to be retained, then it should incorporate (a) modifications to highlight the impact of labour market influences on post-compulsory participation and (b) an adjustment to account for the impact of part-time students on the level of enrolments.

6. Tasmania said that it supported the use of 'age participation rates as a measure of the demand for post-compulsory schooling' for a number of reasons — including its low susceptibility to changing external influences. It did not support use of 'an inferred retention rate' as a measure of post-compulsory schooling as proposed by Victoria. It supported the Commission's intended modelling exercise to test policy and non-policy effects driving the differential demand for post-compulsory enrolments between States. Such an exercise it said should include variables to represent the impact of employment, industry-mix and occupational profile as well as Commonwealth policy influences. It said that it anticipated that at least some of the variables in the Commission's list would interact with each other to produce perverse results.

7. Tasmania suggested that the Commission test the consistency of results obtained from its proposed model and that of Lamb's by comparing it with the standardised participation derived from the application of national average participation rates.

8. Northern Territory said that it was concerned with both retention and participation rates as bases to estimate post-compulsory enrolments. Its preference was for the use of standardised participation rates because of its belief that these provided a better estimate of the number of students in the system and removed the effects of policy. With regard to Lamb's model, it said that the explanatory variables used in the model were potentially highly correlated as exemplified by the correlation between Indigenous people and remote locations and remote locations and socio-economic status. It said that Lamb's report failed to discuss testing for these inter-relationships. It said that participation rate should be used as the dependent variable. The Northern Territory thought that it would be sufficient to include variables for socio-economic status and urbanisation because using more variables would complicate the assessment.

9. The Australian Capital Territory supported using Lamb's retention rate model, presented in Victoria's main submission, as the basis for further research into post-compulsory education enrolments. It said that the approach which Western Australia had adopted — namely, analysing individual schools to support its contention that State policies drive post compulsory education — was not as robust as the approach adopted in Lamb's retention rate study or the study that the ACT commissioned from NATSEM. It concluded that any method should begin with actual rates, with minimal adjustments to account for State policy differences if considered necessary. The ACT did not comment on the second Lamb model.