SUPPORT DOCUMENT

Unmet student demand for tertiary education: Support document

MARGARET GILES

MICHAEL MCCLURE

A MICHAEL DOCKERY

This document was produced by the authors based on their research for the report *Unmet student demand for tertiary education*, and is an added resource for further information. The report is available on NCVER's website: <http://www.ncver.edu.au>

Contents

Contents 2

Tables and figures 3

Appendix 1: Theoretical Model 4

- The Apply (Selection) Model 4
- The Fail (Substantive) Model 5
- Correcting for Selection Bias 5

Appendix 2: SETIT database 6

Appendix 3: UK and US studies 14

Appendix 4: Tables 18

Appendix 5: Selection (Apply) Model Results 21

Appendix 6: Substantive (Fail) Model Results 24

Appendix 7: Corrected Models Results 26

References 27

Tables and figures

Table S1: Unmet and Met Demand Defined by SETIT 8
Table S2: Post-School Study in 2001 9
Table S3: No Post-School Study in 2001 12
Table S4: Post-School Study in 2001 12
Table S5: US and UK studies of demand for further education 14
Table S6: Age and Application Success 18
Table S7: SEIFA/Income and Application Success 18
Table S8: Prior Education and Application Success 19
Table S9: Sensitivity Analysis for Disaggregating Unmet Demand by Sector 20
Table S10: Results of Probit Estimations for Application Submission 21
Table S11: Results of Probit Estimations for Application Failure (Unmet Demand) 24
Table S12: Results of Bivariate Probit Estimations 26

Appendix 1: Theoretical Model

The Apply (Selection) Model

A simple choice theoretic model is to assume that an individual's preferences for applying to study for a post-school qualification can be modelled as a comparison of the utilities yielded under the two actions (studying, not studying)[1]. In this case a person will apply to study if the utility from studying exceeds the utility that would be derived from not studying.

An individual's utility depends on many factors, including income and subjective factors such as attitude to education. Hence, let

$$U_S = U_S(Y_S, Z) \quad \textbf{(1)}$$

$$U_{NS} = U_{NS}(Y_{NS}, Z) \quad \textbf{(2)}$$

where U_S and U_{NS} are the individual's utilities when they apply to study (S) and do not apply to study (NS), respectively. Y is the individual's level of income. For the purpose of this model, this can be taken to be the perceived level of income. While the individual's unearned and earned incomes might be the same in the two states, the perceived level of income might differ. For example, an individual thinking about applying to study may consider that their income on completion of their studies (net of the costs of studying) will be higher than income without the desired qualification. This would suggest that $Y_S > Y_{NS}$. Another individual may consider the opportunity costs of studying too high. That is, future income may be insufficient to outweigh the foregone income whilst studying and other consumable costs. In this case $Y_S < Y_{NS}$. Z represents the vector of personal characteristics that might affect the utilities via their impact on attitudes to education (age, gender, English proficiency, previous education, current occupation, etc).

The decision to apply to study will therefore be based on the magnitude of y^* where $y^* = U_S - U_{NS}$, the individual's underlying propensity to apply to study. If $y^* \geq 0$, then the individual will apply to study. If $y^* < 0$, then the individual will not apply.

The difference in utilities can be expressed as:

$$y^* = y^*(Y_S, Y_{NS}, Z). \quad \textbf{(3)}$$

The empirical representation of this model requires a specific functional form to be selected for y^*. One possibility is:

$$y^* = \alpha_0 + \alpha_1(Y_S - Y_{NS}) + \alpha_2 Z + \varepsilon \quad \textbf{(4)}$$

where α_0 is a constant, α_1 is a scalar parameter with $\alpha_1 > 0$, α_2 is a vector of parameters that link the individual's characteristics to the likelihood of applying to study, and ε captures unmeasured components of the utility difference.

To implement this approach, however, requires information on the income difference in the alternatives of applying to study and not applying. Clearly the individual's unearned and labour market incomes will drop out of the difference term $(Y_S - Y_{NS})$. Hence this term can be replaced by D, which represents the perceived difference in income between individuals who

[1] This model is based on Farber (1983).

pursue further education and those who don't. There are no values for D in SETIT. Such values would in fact be quite difficult to obtain, as they would depend on the personal rates of return to different qualifications, and the probabilities of the events happening. These events include successfully completing studies and finding a higher paid job on completion of studies.

It would be expected, however, that D will depend on the cost of the course and also on the level and field of study of the course. For example, more costly courses are less likely to be applied for. Fields of study that are generalist and have a wider market are more likely to be applied for. Hence the difference term $Y_S - Y_{NS} = D$ can be written as:

$$D = D(C, L, F) \quad \textbf{(5)}$$

where

C = opportunity cost of the course, and

L = level of the course

F = field of study of the course

Through substitution y^* can be written as:

$$y^* = \alpha_0 + \alpha_1[D(C, L, F)] + \alpha_2 Z + \varepsilon \quad \textbf{(6)}$$

However, y^* is unobserved. What is observed is an individual either applying to study ($Y_1 = 1$ if $y^* \geq 0$) or not applying to study ($Y_1 = 0$ if $y^* < 0$). Hence the estimating equation for applying to study is:

$$Y_1 = \alpha_0 + \alpha_1 D + \alpha_2 Z + \varepsilon \quad \textbf{(7)}$$

where Y_1 is a dichotomous variable, D is a vector of course-related variables and Z is a vector of personal characteristics.

The Fail (Substantive) Model

The empirical representation of this model also requires a specific functional form to be selected for Y_2. One possibility is:

$$Y_2 = \alpha_0 + \alpha_1 D + \alpha_2 Z + \varepsilon \quad \textbf{(8)}$$

where Y_2 is a dichotomous variable, D is a vector of course-related variables and Z is a vector of personal characteristics. The components of D and Z in **(8)** will be similar to the components of **(7)**. At least one element of D or Z in **(8)** is not in D or Z in **(7)** (Maddala, 1983).

Equation **(7)** is the estimable substantive equation and equation **(8)** is the estimable selection equation.

Correcting for Selection Bias

Consistent estimates using the bivariate probit procedure are obtained by maximising the log likelihood function:

$$+ y(1-T)\log\Phi[-Zd, Xb; -\rho] + (1-y)(1-T)\log\Phi[-Zd, -Xb; \rho]) \quad \textbf{(9)}$$

In the presence of selectivity bias, the covariance of the error terms, rho, will be non-zero.

Appendix 2: SETIT database

The data analyses in this paper are conducted using the 2001 Survey of Education, Training and Information Technology (SETIT) micro data. This survey was conducted in both urban and rural areas in all States and Territories between April and August 2001. The scope of the survey was all persons aged 15 to 64 years who were usual residents of private dwellings. The questions (asked during face-to-face interviews) included socio-demographic characteristics (such as age and gender), employment characteristics (such as occupation and labour force status), educational qualifications, access to and use of education and training opportunities, and access to and use of information technology (IT). The working database contains 676 variables (some related to multiple response characteristics) and 24,377 records.

SETIT differs from the four yearly Survey of Education and Work (SEW), formerly Transition to Work, in a number of ways. SETIT is a one-off survey conducted over three months in 2001; SEW is conducted as a supplement to the May Monthly Labour Force Survey. SETIT is conducted with all persons in the sampled households whereas SEW is completed by one person on behalf of all people in the household. SEW has 35 times the number of records than SETIT – the May 2001 SEW has 840,400 records compared with SETIT which has 24,377 records. However, SETIT has many more variables than SEW. Finally, SEW is confined to people aged 15 – 64 years who applied to attend an educational institution other than a school in 2001, whereas SETIT includes but is not restricted to this group.

SETIT can be partitioned into 37 sub-populations that are not mutually exclusive. The population groups of interest for this project are: all persons (Population 1); persons not attending school (Population 19); persons who enrolled to study for a non-school qualification in 2000 and continued in 2001 (Population 21) and persons who enrolled to study in 2001 (Population 22) respectively; and persons who did not enrol to study for a non-school qualification in 2001 (Population 26).

The ABS requires that weights be used when analysing these data. The appropriate weights are the inverse of the probability of the inclusion of each respondent in the survey. These are provided in the dataset. Thus commentary on the data, unless otherwise stated[2], refers to a population of 12,870,603 persons aged 15 to 64 years rather than a sample of 24,377 persons. This population corresponds roughly to the size of the Australian population aged 15 years and over in the 2001 Census (N = 14,856,774). Seasonal effects on survey responses are not excluded from the data. Also, there may be small variations between the labour force estimates in SETIT and the corresponding estimates from the Monthly Labour Force Surveys conducted during April to August 2001.

SETIT contains three variables under the general heading of unmet demand. The variable for whether a person applied to enrol in a course of study in 2001, ENROLBAD, has four categories. These are: 'did not apply', 'applied and was unsuccessful', 'applied and was successful or deferred', and 'applied with unknown outcome'.

2 The multivariate analyses necessarily use the unweighted observations because the weighted number of cases introduces large sample bias. That is, very large samples produce extremely small standard errors. Hence significance tests tend to show significance at the 1% level. This is not a helpful result when trying to tease out the various and important influences on application submission or failure. Moreover, the purpose of the weights for reducing the identification of individuals is not violated by the use of unweighted observations in the regressions.

A second variable for the main reason for unsuccessful application, STUDNOGO, has seven categories. These include unsuccessful application due to: 'the course being full', 'the course being cancelled', 'being ineligible or having low entrance scores', 'having applied too late', and 'other reasons'. Two further categories are ' application successful or deferred or outcome unknown', and 'did not apply to enrol'.

The third unmet demand variable relates to whether a person unsuccessfully applied to enrol in a course of study or the course was not in preferred field or institution. This variable, USAPPRE, has five categories. These are: 'unsuccessful application', 'successful application but not preferred field and/or institution', 'successful application with preferred field of study and institution', 'no application' and application with either 'outcome unknown' or 'successful or deferred outcome'. Here, institutional preference is treated in the same way as field of study preference. As mentioned earlier, whether or not the institution at which the individual is enrolled is preferred is not material to the definitions used in this project. For the enumeration of demand, only field of study preference is going to influence whether a particular enrolment represents demand that is met or unmet.

The final USAPPRE category of 'application successful or deferred, or outcome unknown' can be disaggregated using the application variable, ENROLBAD, into 'application successful or deferred' and 'outcome unknown'. The former category applies to persons who are not enrolled in post school study but whose applications were successful. There a number of reasons for this. First, if an application is successful but the applicant has not pursued post-school study, then this may or may not represent unmet demand. Reasons for not following through from application to enrolment might suggest course/institution reasons such as timetable clashes (with work or other commitments). This could represent a form of unmet demand. If the reason is change of mind or family/work commitments[3] then this would not necessarily represent unmet demand. A deferred application could be considered met demand. Again, the reason for this may shed light on whether demand is being met despite there not being an enrolment. 'Outcome unknown' could mean a successful (enrolled or not, or deferred) or unsuccessful application. The data do not allow the separation of these three sub-categories.

Table S1 shows how these variables can be used to derive met and unmet demand, but the result is inconclusive because successful enrolment in non-preferred field (a potential component of unmet demand) is not isolated from non-preferred institution (a component of met demand). An alternative approach, which isolates successful enrolment in non-preferred field from non-preferred institution, using variable CSPREFFS, is shown in Table S2.

3 It may be that an individual with this reason for not enrolling could enrol if adequate day-care or study leave was available. In a sense, this situation could be labelled unmet demand.

Table S1: Unmet and Met Demand Defined by SETIT

<table>
<tr><td rowspan="3">Did not apply

(83.2%)</td><td colspan="8">Did apply

(16.8%)</td></tr>
<tr><td colspan="2">Successful/enrolled

(14.9%)</td><td colspan="5">Unsuccessful

Unmet demand

(0.6%)</td><td rowspan="2">Successful/deferred but not enrolled

(1.0%)</td><td rowspan="2">Outcome unknown and not enrolled

(0.3%)</td></tr>
<tr><td>Preferred field of study and institution

Met demand

(12.9%)</td><td>Field of study or institution not preferred

Met/Unmet demand

(2.0%)</td><td>Course full

Unmet demand

(0.2%)</td><td>Course cancelled

Unmet demand

(0.1%)</td><td>Not eligible

(0.1%)</td><td>Applied too late

(0.1%)</td><td>Else

Met/Unmet demand

(0.2%)</td></tr>
</table>

Note: Includes persons not in school only (n = 12,183,684).

Table S2: Post-School Study in 2001

<table>
<tr><td colspan="6">Not at school</td></tr>
<tr><td colspan="3">No post-school study
(85.1%)</td><td colspan="2">In Vocational Education and Training
(7.7%)</td><td colspan="2">In Higher Education
(6.8%)</td><td colspan="2">In post-school, level not determined
(0.5%)</td></tr>
<tr><td colspan="2">Applied
(1.9%)</td><td>Did not apply
(83.2%)</td><td>Not preferred
Unmet demand
(0.8%)</td><td>Preferred
Met demand
(6.8%)</td><td>Not preferred
Unmet demand
(0.3%)</td><td>Preferred
Met demand
(6.5%)</td><td>Not preferred
Unmet demand
(0.1%)</td><td>Preferred
Met demand
(0.4%)</td></tr>
<tr><td>Unsuccessful
Unmet demand
(0.6%)</td><td>Outcome unknown
(0.3%)</td><td>Successful or deferred
(1.0%)</td><td colspan="6"></td></tr>
</table>

Note: n = 12,183,684

For the purposes of this study, only preferred field of study, as per Table S2, will be used to partition demand. Notwithstanding this choice, there are some remaining issues regarding enumerating met and unmet demand from within SETIT.

Individuals who are not enrolled but who had successful or deferred applications could be added to the met demand figure. However, if enrolment is, as discussed earlier, the key to differentiating met and unmet demand then clearly this group should be excluded from met demand. Whether this group should then be enumerated in the unmet demand category is equivocal. As defined earlier, unmet demand reflects individuals who are frustrated in their attempts to enrol in their preferred course; that is their applications are unsuccessful in some way. This may not be the case with individuals who have successful or deferred applications for study in 2001 but who chose not to enrol at that time. Of these, 5.6% were not enrolled for course- or qualification-related reasons; 94.4% were not enrolled for other (work, personal or other) reasons.

To ensure that total demand is fully enumerated, two other variables in SETIT have been examined. These are the variable for whether a person wanted to study for an(other) educational qualification in the last 12 months (up to the survey date), and the main reason for wanting but not being able to study. Tables S3 and S4 use these variables and the variables from Tables S1 and S2 to further extract further cases of student demand, both met and unmet. Figure S1 summarises the derivation of unmet demand from the weighted observations.

Figure S1: Derivation of Total Demand

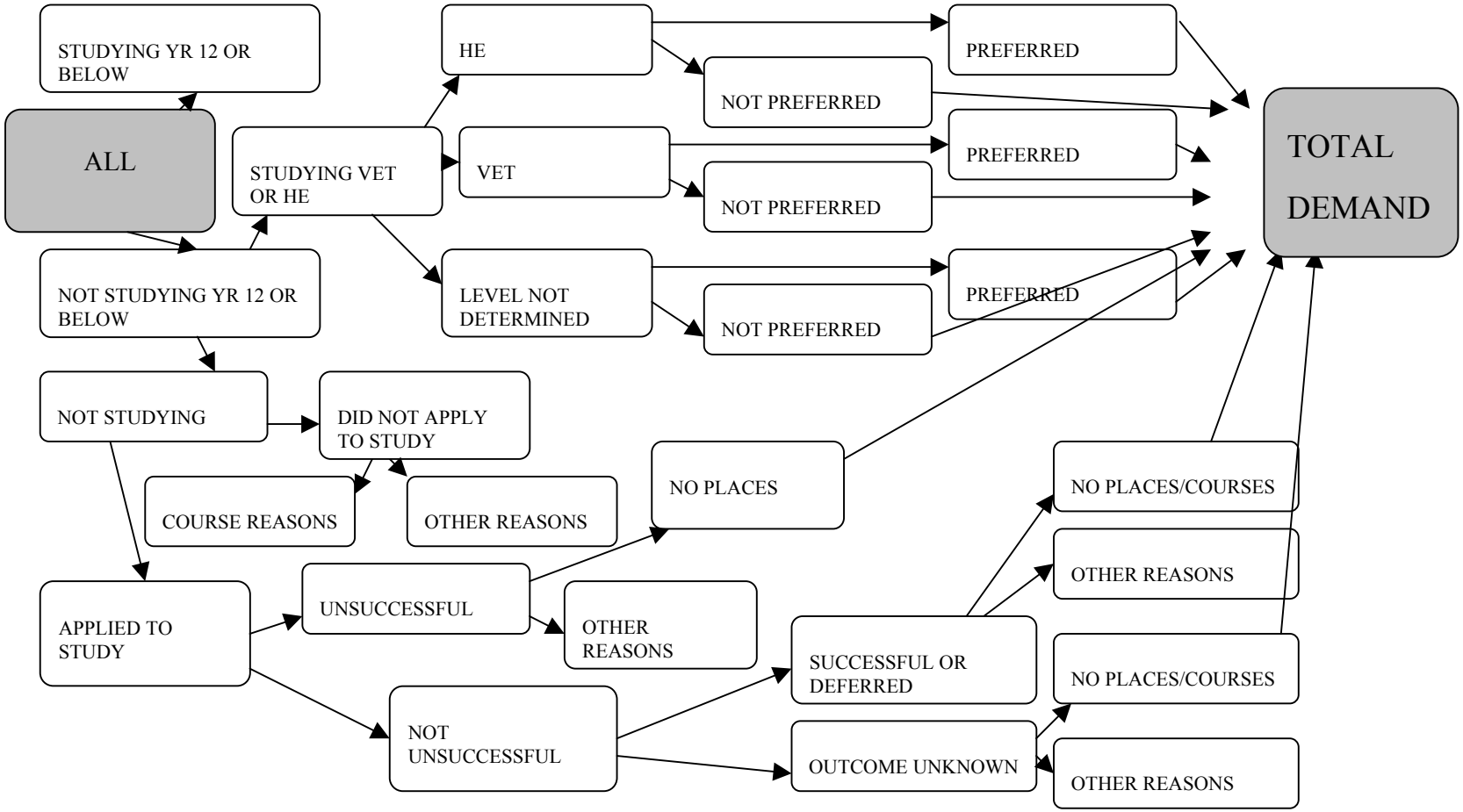

Table S3: No Post-School Study in 2001

No post-school study 10,366,415 = 85.1%									
Wanted to study 2,166,430 = 20.9%								**Did not want to study 8,199,984 = 79.1%**	
No places/suitable courses 312,830 = 14.4%				**Other reasons 1,853,600 = 85.6%**				**No places/suitable courses 0 = 0%**	**Other reasons 8,199,984 = 100%**
Applied 26,104 = 8.3%			**Did not apply 286,726 = 91.7%**	**Applied 75,420 = 4.1%**			**Did not apply 1,778,180 = 95.9%**		
Application success unknown 3,040 = 11.6%	**Application successful or deferred 6,755 = 25.9%**	**Application unsuccessful 16,309 = 62.5%**		**Application success unknown 12,521 = 16.6%**	**Application successful or deferred 34,661 = 46.0%**	**Application unsuccessful 28,238 = 37.4%**			

Table S4: Post-School Study in 2001

Post-school study 1,817,268 = 14.9%					
Studying with level not determined 58,322 = 0.5%		**In VET 933,551 = 7.7%**		**In higher education 825,395 = 6.8%**	
Preferred field of study 47,806 = 82.0%	**Not preferred field of study 10,515 = 18.0%**	**Preferred field of study 831,148 = 89.0%**	**Not preferred field of study 102,403 = 11.0%**	**Preferred field of study 786,598 = 95.3%**	**Not preferred field of study 38,798 = 4.7%**

The presumption in the data is that definitions of met demand reflect individuals' desires to study in the same way that some of those who are undertaking no post-school study in 2001 may have wanted to study. Two examples can be used to question this presumption. First, for some individuals, study will have been a 'fall-back' option if they failed to get a job or their preferred job. Second, for some individuals, study preferences may have been less important than the need to access an income stream such as Austudy. In both examples, wanting to study is a moot point. Nonetheless these individuals have applied to and been accepted into publicly funded or private institutions for the purpose of post school study. In this sense they are included in estimates of demand. In terms of the SETIT data, these individuals cannot be distinguished from others who might more legitimately be classified as met demand. To the extent that the former group is small relative to the latter group, or that characteristics affecting application failure or the decision to apply are comparable between the two groups, this delineation may not matter. If there are statistically significant differences, then the results of the analyses will be biased to some unknown degree.

Appendix 3: UK and US studies

Table S5: US and UK studies of demand for further education

Factors affecting demand[1]	Author(s)	Where	Year of survey	Sample
Ability (maths and reading test scores)	Hagy and Staniec (2002)	US	1988 - 1994	10,222 year 8 cohort from 1988; National Educational Longitudinal Study
Ability difference	Kohn, Manski and Mundel (1972; 1973)[2]	US	1966	3015 high school seniors
Academic ability	Radner and Miller (1970)[2]	US	1966	1402 high school seniors
Academic attainment	McVicar and Rice (2001)	UK	1955 - 1994	various
Academic selectivity	Radner and Miller (1970)[2]	US	1966	1402 high school seniors
Age	Heckman and Smith (2003)	US	1987 – 1989	National Job Training Partnership Act (NJTPA) participants
Average hourly earnings of production workers	Corrazzini, Dugan and Grabowski (1972)[2]	National and Massachusetts, US	1963	Tenth graders in 1960 who enrolled in college in 1963
Average real household income	McVicar and Rice (2001)	UK	1955 - 1994	various
Average student ability	Kohn, Manski and Mundel (1972; 1973)[2]	US	1966	3015 high school seniors
Born overseas or at least one parent born overseas	Hagy and Staniec (2002)	US	1988 - 1994	10,222 year 8 cohort from 1988; National Educational Longitudinal Study
Breadth of offering	Kohn, Manski and Mundel (1972; 1973)[2]	US	1966	3015 high school seniors
Coeducational	Kohn, Manski and Mundel (1972; 1973)[2]	US	1966	3015 high school seniors

College revenue per student	Kohn, Manski and Mundel (1972; 1973)[2]	US	1966	3015 high school seniors
College type	Kohn, Manski and Mundel (1972; 1973)[2]	US	1966	3015 high school seniors
Cost of college	Kohn, Manski and Mundel (1972; 1973)[2]	US	1966	3015 high school seniors
Costs of alternative post-secondary options	Radner and Miller (1970)[2]	US	1966	1402 high school seniors
CPI	Campbell and Siegel (1967)[2]	US	1927 – 1963 (9 yrs)	Four-year institutions
Dormitory capacity	Kohn, Manski and Mundel (1972; 1973)[2]	US	1966	3015 high school seniors
Education	Heckman and Smith (2003)	US	1987 – 1989	National Job Training Partnership Act (NJTPA) participants
English proficiency	Heckman and Smith (2003)	US	1987 – 1989	National Job Training Partnership Act (NJTPA) participants
Family income	Heckman and Smith (2003)	US	1987 – 1989	National Job Training Partnership Act (NJTPA) participants
Family income	Kohn, Manski and Mundel (1972; 1973)[2]	US	1966	3015 high school seniors
Family income	Radner and Miller (1970)[2]	US	1966	1402 high school seniors
Female	McVicar and Rice (2001)	UK	1955 - 1994	various
Foreign born black and Hispanics	Vernez and Abrahamse (1996)[3]	US	1980	High School and Beyond sophomore and senior classes data
Foreign born Hispanics	Ganderton and Santos (1995)[3]	US	1980	High School and Beyond senior class data
Foreign born whites and Asians	Vernez and Abrahamse (1996)[3]	US	1980	High School and Beyond sophomore and senior classes data
High school location	Hagy and Staniec (2002)	US	1988 - 1994	10,222 year 8 cohort from 1988; National Educational Longitudinal Study
Income	Hagy and Staniec (2002)	US	1988 - 1994	10,222 year 8 cohort from 1988; National Educational Longitudinal Study
Income	Hoenack (1967)[2]	California, US	1967-68	Eligible high school graduates
Income	Hoenack, Weiler and Orvis (1973)[2]	Minnesota, US	1948 – 1972	Eligible high school graduates
Income – low, middle, high	Jackson and Weathersby (1975)	Review of other studies	Various	Various

Individual's own subjective probability of graduating	Hilmer (1998)	US	1980	High School and Beyond sophomore and senior classes data
Marital status	Heckman and Smith (2003)	US	1987 – 1989	National Job Training Partnership Act (NJTPA) participants
Mother born overseas	Fligstein and Fernandez (1985)[3]	US	1979	National Longitudinal Survey
Number of siblings	Hagy and Staniec (2002)	US	1988 - 1994	10,222 year 8 cohort from 1988; National Educational Longitudinal Study
Parent's educational attainment	Hagy and Staniec (2002)	US	1988 - 1994	10,222 year 8 cohort from 1988; National Educational Longitudinal Study
Parent's occupational status	Hagy and Staniec (2002)	US	1988 - 1994	10,222 year 8 cohort from 1988; National Educational Longitudinal Study
Parental education	Burnhill, Garner and McPherson (1990)	Scotland	1977 – 1985	Scottish School Leavers Survey – biennial postal surveys
Parental social class/occupation	Burnhill, Garner and McPherson (1990)	Scotland	1977 – 1985	Scottish School Leavers Survey – biennial postal surveys
Participation in government transfer programs	Heckman and Smith (2003)	US	1987 – 1989	National Job Training Partnership Act (NJTPA) participants
Paternal education	Corrazzini, Dugan and Grabowski (1972)	National and Massachusetts, US	1963	Tenth graders in 1960 who enrolled in college in 1963
Percentage of high school class enrolled in college preparation classes	Hagy and Staniec (2002)	US	1988 - 1994	10,222 year 8 cohort from 1988; National Educational Longitudinal Study
Private/public school	Hagy and Staniec (2002)	US	1988 - 1994	10,222 year 8 cohort from 1988; National Educational Longitudinal Study
Provision of higher education	McVicar and Rice (2001)	UK	1955 - 1994	various
Provision of youth training	McVicar and Rice (2001)	UK	1955 - 1994	various
Quality (college combined SAT score)	Spies (1973)[2]	US		Applications to college
Race	Heckman and Smith[1] (2003)	US	1987 – 1989	National Job Training Partnership Act (NJTPA) participants
Race/ethnicity (black, white, Hispanic, Asian)	Hagy and Staniec (2002)	US	1988 - 1994	10,222 year 8 cohort from 1988; National Educational Longitudinal Study
Recent labour force status	Heckman and Smith (2003)	US	1987 – 1989	National Job Training Partnership Act (NJTPA) participants
Relative earnings	McVicar and Rice (2001)	UK	1955 - 1994	various

Sex	Hagy and Staniec (2002)	US	1988 - 1994	10,222 year 8 cohort from 1988; National Educational Longitudinal Study
Social class	McVicar and Rice (2001)	UK	1955 - 1994	various
Socioeconomic quartile	Corrazzini, Dugan and Grabowski (1972)[2]	National and Massachusetts, US	1963	Tenth graders in 1960 who enrolled in college in 1963
Student country of origin	Fligstein and Fernandez (1985)[3]	US	1979	National Longitudinal Survey
Type of institution	Hagy and Staniec (2002)	US	1988 - 1994	10,222 year 8 cohort from 1988; National Educational Longitudinal Study
Youth unemployment	McVicar and Rice (2001)	UK	1955 - 1994	various

Notes:

1. There may be other factors attributed to the studies that are included with other studies.
2. These studies are summarised in Jackson and Weathersby (1975).
3. These studies are summarised in Hagy and Staniec (2002).

Appendix 4: Tables

Table S6: Age and Application Success

Age group	% of sample	% successful	χ^2	p
15 – 19 years	19.2	82.1	0.894	0.344
20 – 24 years	23.4	86.7	25.071	0.000
25 – 29 years	13.2	80.7	0.011	0.918
30 – 34 years	11.1	78.6	1.523	0.217
35 – 39 years	10.1	76.5	5.213	0.022
40 – 44 years	8.8	74.8	8.715	0.003
45 – 49 years	6.5	80.3	0.046	0.829
50 – 54 years	4.5	76.9	1.765	0.184
55 – 59 years	2.3	75.3	1.736	0.188
60 – 64 years	0.9	79.4	0.046	0.830
All	100.0	80.9		

Note: n = 3,766
Source: SETIT 2001

Table S6 shows that the age groups 20 to 24 years and 40 to 44 years have chi square values that are significant at the 1% level and the age group 35 to 39 years has a chi square value that is significant at the 5% level. For the analysis in Appendix 6, the two older age groups are combined into a single group 35 to 44 years. In the probit estimation for the model of application failure, it is expected that the coefficient of the 20 to 24 year age group will be positive and the coefficient of the 35 to 44 year age group will be negative.

Table S7: SEIFA/Income and Application Success

Characteristic	% of sample	% successful	χ^2	p
SEIFA – 1st quartile	13.9	76.6	7.086	0.008
SEIFA – 2nd quartile	16.9	74.5	20.106	0.000
SEIFA – 3rd quartile	20.1	82.7	2.131	0.144
SEIFA – 4th quartile	22.8	82.5	1.917	0.166
SEIFA – 5th quartile	26.2	84.4	10.617	0.001
Income below median	40.8	82.5	4.809	0.028
Median income	2.3	80.0	0.041	0.840
Income above median	21.7	80.4	0.160	0.689
Income not known	35.2	79.2	3.415	0.065
All	100.0	80.9		

Note: n = 3,776
Source: SETIT 2001

Table S7 shows that there are significant differences between the success or not of applications for post school study for those respondents in the 1st, 2nd and 5th quartiles of SEIFA (less likely, less likely and more likely to have successful applications for post school study respectively) and for those whose income is below the median (more likely to have a successful application). These results appear contradictory. However, SEIFA is a composite of four indices within which income is only one of many components. These four indices are disadvantage (variables include income, educational attainment and unemployment), advantage/disadvantage (variables include the proportion of families with high incomes, tertiary education, etc), economic resources (variables include income, expenditure and assets) and education and occupation (variables include proportions of people with higher qualifications or employed in skilled occupations) (Australian Bureau of Statistics, 2003). So it is not unlikely that an applicant with low income (due to age for example) may reside in an area in the top SEIFA quintile.

Table S8: Prior Education and Application Success

Highest level of education attained	% of sample	% successful	χ^2	p
Postgraduate	6.8	78.3	1.171	0.279
Bachelor	15.1	83.3	2.600	0.107
Diploma	8.9	82.7	0.799	0.372
Certificate	13.7	77.0	5.796	0.016
Year 12	35.9	85.0	23.159	0.000
Year 11	5.8	72.1	11.384	0.001
Year 10 or below	12.8	74.5	14.283	0.000
No schooling	0.0	50.0	1.230	0.267
Level not determined	1.0	83.3	0.144	0.704
Not Year 12	31.9	74.6	41.363	0.704
All	100.0	80.9		

Note: n = 3,766
Source: SETIT 2001

Respondents who have not completed Year 12 may be less likely to have successful applications due to a lack of prerequisites. This cannot be tested within SETIT. However, it appears that unsuccessful applicants who have completed Year 12 are as likely as those without Year 12 completion to have reasons for not studying including lack of information, no suitable courses, no places available and not being offered a place[4]. The coefficient for the variable 'Not Year 12' is expected to be negative.

A finer level of disaggregation of prior education is obtained by looking at the ASCED range of highest level of education attained. This range is shown in Table S8 together with the chi square results of testing the correlation between these categories of prior education and application success. Also included in the table is the result for testing the binary variable 'Not Year 12'.

Table S9 provides a sensitivity analysis for disaggregating unmet demand for post school education between higher education (HE) and vocational education and training (VET). Discussion of this table can be found in the conclusions in the main report.

4 Of unsuccessful applicants without Year 12 completion, 4.8% cited course-related reasons. Of unsuccessful applicants with Year 12 completions, 3.6% cited course-related reasons. These differences are not significant (χ^2 = 2.199, p = 0.138).

Table S9: Sensitivity Analysis for Disaggregating Unmet Demand by Sector

Sector	Post school enrolments in 2001[2] (%)	Highest level of educational attainment in 2001[3] (%)	Unmet Demand[4]		
			Same as post school enrolments ('000)	VET as one third ('000)	VET as 40.7%[5] ('000)
	(1)	(2)	(3)	(4)	(5)
HE	45.6	26.2	10.5	15.4	13.7
VET	51.7	17.9	11.9	7.7	9.4
Don't know	3.4	0	0.4	0.0	0.0
Else	0	55.9	0.0	0.0	0.0
Total[1]	100.0	100.0	23.1	23.1	23.1

Notes:

1. May not sum due to rounding.
2. From SETIT weighted observations.
3. From SETIT weighted observations.
4. Using unmet demand of 23,064 (1.3% of total demand).
5. 40.7% = (17.9/(17.9 + 26.2)) x 100. (Using weighted observations).

Appendix 5: Selection (Apply) Model Results

For the models shown in Table S10, the estimation technique used probit as the dependent variable, APPLY, is dichotomous. The sample was restricted to respondents who were not, at the time of the survey, studying at school. Five models with differing combinations of factors were estimated. All factors are constructed as binary variables.

Table S10: Results of Probit Estimations for Application Submission

Variables	Model 1	Model 2	Model 3	Model 4	Model 5
15 to 19 years	1.882*	1.874*	1.891*	1.805*	n.a.
20 to 24 years	1.384*	1.377*	1.401*	1.348*	n.a.
25 to 29 years	0.976*	0.969*	1.009*	1.003*	n.a.
30 to 34 years	0.863*	0.856*	0.923*	0.931*	n.a.
35 to 39 years	0.770*	0.766*	0.852*	0.859*	n.a.
40 to 44 years	0.690*	0.684*	0.764*	0.776*	n.a.
45 to 49 years	0.589*	0.585*	0.628*	0.634*	n.a.
50 to 54 years	0.421*	0.418*	0.433*	0.449*	n.a.
55 to 59 years	0.267*	0.265*	0.270*	0.281*	n.a.
Female	0.092*	0.092*	0.098*	0.052**	-0.001
SEIFA – second quintile	0.072**	n.a.	n.a.	n.a.	n.a.
SEIFA – third quintile	0.040	n.a.	n.a.	n.a.	n.a.
SEIFA – fourth quintile	0.078**	n.a.	n.a.	n.a.	n.a.
SEIFA – fifth quintile	0.077**	n.a.	n.a.	n.a.	n.a.
Not finished Year 12	-0.398*	-0.406*	-0.395*	-0.410*	-0.549*
Father born overseas	-0.007	-0.006	-0.007	-0.006	0.016
Mother born overseas	-0.056	0.055	-0.056	-0.053	-0.011
Computer at home	0.491*	0.499*	0.501*	0.513*	0.544*
Born overseas	0.137*	0.137*	0.138*	0.132*	-0.039
Not married	0.404*	0.402*	0.361*	0.346*	0.667*
Not working	0.144*	0.138*	0.169*	0.342*	0.423*
ESL	-0.024	0.028	-0.032	-0.035	-0.043
With children	n.a.	n.a.	-0.140*	-0.147*	-0.064**
With disability	n.a.	n.a.	-0.081*	-0.087*	-0.208*
Income below the median	n.a.	n.a.	n.a.	0.348*	0.452*
Income above the median	n.a.	n.a.	n.a.	0.146**	0.029
Income not known or zero	n.a.	n.a.	n.a.	0.052	-0.072
Constant	-2.333*	-2.271*	-2.247*	-2.417*	1.545*
Log likelihood	-8286.807	8289.915	8274.132	8232.248	8737.842

Note: n = 23,048 (unweighted SETIT observations)

The coefficients in the probit model are not directly interpretable. However, a positive coefficient indicates an increase the probability of applying. A negative coefficient indicates a decrease in the probability of applying.

In Model 1, the independent variables are those that have been shown by other studies to be important factors in post school participation and course and/or institution choices. There are nine age variables with a benchmark category of 60 to 64 years. All coefficients of the age variables are positive as hypothesised. They are also large and significant at the 1% significance level. Thus, younger people, particularly those aged 15 to 19 years, are much more likely than older people to apply for post school study. This is a substantial impact and conforms to the policy and practice of encouraging continued education and training for school leavers. As expected, the magnitude of the age coefficients diminishes as age increases.

In this model, females are more likely to apply than males, persons without Year 12 completions are less likely to apply than those who have completed Year 12, persons with access to computers at home are more likely to apply than those without home access to computers, persons who not working are more likely to apply than those who are working, and unmarried people are more likely to apply than married people. These results are consistent with the hypothesised signs of the estimates. A further variable with a coefficient that is significant at the 1% level is 'Born overseas'. The sign here is positive suggesting that the influences promoting education for migrants outweigh the difficulties that may arise, for example, from language and cultural differences.

In Model 2, the SEIFA variables are excluded. SEIFA itself has a number of dimensions including income. It was felt that these may have conflicting influences on decisions to apply for further education such that these SEIFA variables have no aggregate importance. The coefficients of the remaining variables are similar in sign, magnitude and significance to the like variables in Model 1. The log likelihood of Models 1 and 2 is about the same.

Model 3 is similar to Model 2 with the inclusion of a further two variables. These are parenting status, and having a disability. The age variables have coefficients that are large, positive and significant at the 1% level. The gender regressor is positive and significant. Of the remaining variables with significant coefficients, 'Not finished Year 12', 'With disability' and 'With children' have negative signs and 'Computer at home', 'Born overseas', and 'Not married' have positive signs. Excluding 'Born overseas' for which the hypothesised effects were inconclusive with regard to sign, these results support the hypotheses.

Income variables were added to the Model 3 variables in the estimation of Model 4. In this model, the age variables have coefficients which are significant and magnitudes that are large. The coefficient of gender is significant only at the 5% level and its magnitude is very small. The labour force variable tends to be correlated with the income variable in that an individual who is not working or not in the labour force is unlikely to have an income. Conversely, an individual who is working will have an income. This suggests that for those who are working, the income variables are picking up most of the effect on the decision to apply; for those not working the labour force variable is picking up most of the effect. The magnitude, sign and significance of the other variables in Model 4 are the same as for Model 3. People on incomes below the median level are slightly more likely to apply for post school study than persons earning the median level income.

In Model 5, the effects of variables are examined without controlling for age. A comparison of estimates for Models 4 and 5 shows that, for all non-age variables, the magnitudes, signs and/or significance levels are different. For example, the regressor for the gender variable 'Female' is small, positive and significant at the 5% level in Model 4 and is very small, negative and insignificant in Model 5. This suggests some interactions between age and gender in the decision

to apply for post school study[5]. The regressor for the ethnicity variable 'Born overseas' is small, positive and significant in Model 4 and much smaller, negative and insignificant in Model 5. To some extent the estimates in Model 5 are picking up the direct effect of each respective variable on the decision to apply for post school study plus indirect effects that should be attributed to the omitted age variables (Gujarati, 1995).

5 Using the unweighted sample however, the correlation between gender and age for applicants is not significant (χ^2 = 9.359, df = 9, p = 0.405).

Appendix 6: Substantive (Fail) Model Results

Bivariate tests were conducted for the variables shown in Table S10. A chi square test showed a correlation between application success and whether the applicant's father was born overseas or in Australia ($\chi^2 = 3.837$, p = 0.050). About 82% of applications where the applicant's father was born overseas are successful compared with 80.9% of all applications. The coefficient for this variable is expected to be negative. Neither 'Self born overseas' or 'Mother born overseas' showed any significant correlation between application success and place of birth.

A slightly higher proportion of applicants who are not married have successful applications ($\chi^2 = 6.020$, p = 0.014). The coefficient of this variable is expected to be negative. Respondents with English as a second language are slightly more likely to have successful applications. However this result is insignificant ($\chi^2 = 1.808$, p = 0.179). Fewer respondents with English as a second language apply for post school study (16.0%) compared with the total of respondents (16.4%). However this result is not significant ($\chi^2 = 0.480$, p = 0.488). A reasonable hypothesis is that many ESL applicants are choosing less popular courses (for example, language courses) or are enrolling in pre-course self funded ESL courses for which applications are unlikely to be knocked back.

Reference was made earlier to problems that people with disabilities continue to have when attempting to access services, including education. It is possible that application processes still discriminate against disabled people for many reasons. On balance, it is expected that the coefficient for this variable will be positive.

For the models shown in Table S11, the estimation technique used probit as the dependent variable, FAIL, is dichotomous. The sample was restricted to respondents who were not, at the time of the survey, studying at school and who had applied for post school study in 2001. Three models of unmet demand (failed applications) with differing combinations of factors were estimated. All factors are constructed as binary variables.

Table S11: Results of Probit Estimations for Application Failure (Unmet Demand)

Variables	Model 1	Model 2	Model 3
15 to 19 years	n.a.	0.022	n.a.
20 to 24 years	-0.234*	-0.181	-0.229*
25 to 29 years	n.a.	0.035	n.a.
30 to 34 years	n.a.	0.097	n.a.
35 to 39 years	n.a.	0.166	n.a.
40 to 44 years	n.a.	0.247	n.a.
45 to 49 years	n.a.	0.040	n.a.
50 to 54 years	n.a.	0.147	n.a.
55 to 59 years	n.a.	0.190	n.a.
35 to 44 years	0.138**	n.a.	0.141**
SEIFA 1st quintile	0.056	n.a.	n.a.
SEIFA 5th quintile	-0.075	n.a.	n.a.

Highest level of education attained – post graduate	n.a.	0.132	n.a.
Highest level of education attained – bachelor	n.a.	-0.002	n.a.
Highest level of education attained – diploma	n.a.	-0.015	n.a.
Highest level of education attained – certificate	0.107	0.178**	n.a.
Highest level of education attained – Year 11	0.259**	0.324*	n.a.
Highest level of education attained – Year 10 or below	0.137**	0.219*	n.a.
No schooling	n.a.	0.722	n.a.
Highest level of education attained – Year 12	-0.057	n.a.	n.a.
Not finished Year 12	n.a.	n.a.	0.193*
Father born overseas	-0.059	n.a.	n.a.
Computer at home	-0.324*	-0.343*	-0.343*
Not married	-0.021	n.a.	n.a.
English as a second language	-0.030	n.a.	n.a.
With disability	0.143**	n.a.	0.151*
Income below the median income group	n.a.	-0.010	n.a.
Income above the median income group	n.a.	0.010	n.a.
Income unknown/not applicable	n.a.	0.036	n.a.
Constant	-0.599*	-0.728**	-0.674*
n (unweighted SETIT observations)	3740	3740	3766
Log Likelihood	-1744.317	-1777.037	-1794.866

The results of the estimation of Model 1, show that applications tend to be less successful for older applicants, applicants from relatively disadvantaged areas, applicants who have not completed year 12, applicants with fathers born in Australia, applicants without computers at home, married applicants and applicants with English as a second language. The probability of failure is lower for younger applicants, applicants from relatively advantaged areas, more educated applicants and applicants without a disability.

Model 2 in Table S11 has been estimated with those applicants where the prior education level is known. That is, applicants with highest level of educational attainment not determined are excluded. The results show that the probability of a failed application is lower for applicants aged 20 to 24 years, for applicants without bachelor degrees or diplomas, for applicants without computers at home and for applicants with incomes above the median level of income.

Model 3 is a more parsimonious model. Applicants aged 20 to 24 years are less likely to fail as are applicants without a computer at home. The probability of failure is higher for applicants without Year 12 completions and for applicants with a disability.

Appendix 7: Corrected Models Results

Table S12 gives the results of the maximum likelihood estimates generated by the bivariate probit procedure where the FAIL model is Model 3 from Table S11 and the APPLY model is Model 4 from Table S10.

Table S12: Results of Bivariate Probit Estimations

VARIABLES	FAIL (Model 3)	APPLY (Model 4)
15 to 19 years	n.a.	1.804*
20 to 24 years	-0.217*	1.348*
25 to 29 years	n.a.	1.003*
30 to 34 years	n.a.	0.931*
35 to 39 years	n.a.	0.859*
40 to 44 years	n.a.	0.776*
45 to 49 years	n.a.	0.634*
50 to 54 years	n.a.	0.448*
55 to 59 years	n.a.	0.280*
35 to 44 years	0.127**	n.a.
Female	n.a.	0.052**
Not finished Year 12	0.176*	-0.410*
Father born overseas	n.a.	-0.005
Mother born overseas	n.a.	-0.054
Computer at home	-0.328*	0.513*
Born overseas	n.a.	0.133*
Not married	n.a.	0.346*
Not working	n.a.	0.341*
English as a second language	n.a.	-0.035
With children	n.a.	-0.148*
With disability	0.145**	-0.087*
Income below the median income group	n.a.	0.348*
Income above the median income group	n.a.	0.146**
Income unknown/not applicable	n.a.	0.051
Constant	-0.728*	-2.416*
Log Likelihood	-10026.88	
Disturbance correlation (rho)	0.041	

The null hypothesis for the selectivity bias correction model is that rho is equal to zero (see Appendix 1). The results show that this cannot be rejected as the estimate of rho is positive, small and insignificant at the 1% level. This suggests that there are no unobserved factors that impact on both models. Specifically, excluding non-applicants from the determination of influences on application failure (unmet demand) does not introduce bias.

References

Australian Bureau of Statistics. (2003, November 5). *Assessing the Needs and Opportunities of Australian Communities*. Retrieved November 12, 2003, from http://www.abs.gov.au/ausstats/abs@.nsf/0/D905CF922F757D6FCA256DD4007F7CDF?Open&Highlight=0,SEIFA

Burnhill, P., Garner, C., & McPherson, A. (1990). Parental Education, Social Class and Entry to Higher Education 1976 - 86. *Journal of Royal Statistical Society A, 153*(Part 2), 233 - 248.

Campbell, R., & Siegel, B. (1967). The Demand for Higher Education in the United States, 1919 - 1964. *American Economic Review, June*, 482 - 494.

Corrazzini, A., Dugan, D., & Grabowski, H. (1972). Determinants and Distributional Aspects of Enrolment in US Higher Education. *Journal of Human Resources, Winter*, 39 - 50.

Farber, H. S. (1983). Worker Preferences for Union Representation. *Research in Labor Economics, Supplement 2*, 171 - 205.

Fligstein, N., & Fernandez, R. (1985). Educational Transitions of Whites and Mexican-Americans. In G. Borjas & M. Tienda (Eds.), *Hispanics in the US Economy* (pp. 161 - 192). Orlando, FL: Academic Press.

Ganderton, P., & Santos, R. (1995). Hispanic College Attendance and Completion: Evidence from the High School and Beyond Surveys. *Economics of Education Review, 14*(1), 35 - 46.

Gujarati, D. N. (1995). *Basic Econometrics* (Third edition ed.). New York: McGraw-Hill, Inc.

Hagy, A. P., & Staniec, J. F. O. (2002). Immigrant Status, Race, and Institutional Choice in Higher Education. *Economics of Education Review, 21*, 381 - 392.

Heckman, J. J., & Smith, J. A. (2003). The Determinants of Participation in a Social Program: Evidence from a Prototypical Job Training Program. *NBER Working Paper Series, 9818*.

Hilmer, M. J. (1998). Post-Secondary Fees and the Decision to Attend a University. *Journal of Public Economics, 67*, 329 - 348.

Hoenack, S. (1967). Private Demand for Higher Education in California. Office of Analytical Studies, University of California.

Hoenack, S., Weiler, W., & Orvis, C. (1973). *Cost-Related Tuition Policies and University Enrolments* (mimeo): Management Information Division, University of Minnesota.

Jackson, G. A., & Weathersby, G. B. (1975). Individual Demand for Higher Education. *Journal of Higher Education, XLVI*(6), 623 - 652.

Kohn, M., Manski, C., & Mundel, D. (1972, December). *A Study of College Choice*. Paper presented at the Econometric Society.

Kohn, M., Manski, C., & Mundel, D. (1973). *An Empirical Investigation of Factors Which Influence College Going Behaviors* (mimeo).

Maddala, G. S. (1983). *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

McVicar, D., & Rice, P. (2001). Participation in Further Education in England and Wales: an Analysis of Post-War Trends. *Oxford Economic Papers, 53*, 47 - 66.

Radner, R., & Miller, L. (1970). Demand and Supply in US Higher Education. *American Economic Review, 60*(2), 326 - 334.

Spies, R. (1973). *The Future of Private Colleges: The Effect of Rising Costs on College Choice*. Princeton: Industrial Relations Section, Princeton University.

Vernez, G., & Abrahamse, A. (1996). *How Immigrants Fare in US Education*. Santa Monica, CA: RAND.