

TECHNICAL PAPER

Measuring the socioeconomic status of Australian youth

PATRICK LIM

SINAN GEMICI

NATIONAL CENTRE FOR VOCATIONAL EDUCATION RESEARCH



Australian Government

Department of Education, Employment
and Workplace Relations



Measuring the socioeconomic status of Australian youth

PATRICK LIM
SINAN GEMICI

NATIONAL CENTRE FOR
VOCATIONAL EDUCATION RESEARCH

The views and opinions expressed in this document are those of the author/
project team and do not necessarily reflect the views of the Australian Government
or state and territory governments.

© National Centre for Vocational Education Research, 2011

With the exception of the cover design, artwork, all logos, and any other material where copyright is owned by a third party, all material presented in this document is provided under a Creative Commons Attribution 3.0 Australia <<http://creativecommons.org/licenses/by/3.0/au/>>.



This document should be attributed as Lim, P & Gemici, S 2011, *Measuring the socioeconomic status of Australian youth*.

The National Centre for Vocational Education Research (NCVER) is an independent body responsible for collecting, managing and analysing, evaluating and communicating research and statistics about vocational education and training (VET).

NCVER's inhouse research and evaluation program undertakes projects which are strategic to the VET sector. These projects are developed and conducted by the NCVER's research staff and are funded by NCVER. This research aims to improve policy and practice in the VET sector.

ISBN 978 1 921809 86 6 web edition

ISBN 978 1 921809 87 3 print edition

TD/TNC 103.22

Published by NCVER

ABN 87 007 967 311

Level 11, 33 King William Street, Adelaide, SA 5000

PO Box 8288 Station Arcade, Adelaide SA 5000, Australia

ph +61 8 8230 8400 fax +61 8 8212 3436

email ncver@ncver.edu.au

<<http://www.ncver.edu.au>>

<<http://www.ncver.edu.au/publications/2355.html>>

Measuring the socioeconomic status of Australian youth

Patrick Lim and Sinan Gemici, NCVER

Developing an accurate measurement of individual socioeconomic status (SES) is important, particularly because of current policy interest in increasing educational participation among those from a low SES background. Typically, we do not have data on all the underlying characteristics that make up the concept of SES. Consequently, SES is usually measured indirectly, such as through the Socio-Economic Indexes for Areas (SEIFA) or through parental occupation or education. The issue is how satisfactory these proxies are in measuring 'true' SES.

Using the 2003 cohort of the Longitudinal Surveys of Australian Youth (LSAY), Lim and Gemici focus on measuring the SES of young people aged 15 to 25 years. The authors create a measure of individual SES that captures the impact of cultural and educational resources, as well as parental education and occupation.

The authors find:

- ✧ SEIFA greatly misclassifies SES at the individual level (almost 40% of individuals are wrongly classified as high or low SES).
- ✧ SEIFA composites result in only a marginal improvement in classification accuracy.
- ✧ SEIFA and SEIFA composites perform reasonably well when reporting participation in higher education at aggregate levels.
- ✧ SEIFA and SEIFA composites perform reasonably well in multivariate modelling, although the relationship between participation and socioeconomic disadvantage is underestimated using SEIFA measures.

The implications are that SEIFA is satisfactory when determining aggregate relationships, but performs very poorly when classifying individuals. This is problematic for programs which direct resources to individuals: use of an area-based measure of SES will result in the misallocation of resources.

Tom Karmel
Managing Director, NCVER

Contents

Tables and figures	6
Introduction	7
Relevant dimensions of SES	8
Parental occupation	8
Parental education	9
Household income and wealth	10
Ancillary dimensions of SES	10
Towards creating a measure of SES	11
Creating a reference measure of SES	12
Considering alternative measures of SES	17
Socio-Economic Indexes for Areas	17
SEIFA composites	18
Testing alternative SES measures	19
Performance of SES measures at the individual level	19
Performance of SES measures at the aggregate level	21
Performance of SES measures in multivariate analysis	22
Conclusion	24
References	25
Appendices	
A	26
B	27

Tables and figures

Tables

1	ISCED classification	9
2	LSAY variables used as a basis for creating an SES reference measure	12
3	Initial latent-class factor analysis	13
4	Loadings for the four-factor model (varimax rotation)	13
5	Modified latent-class factor analysis model	14
6	Loadings for the three-factor model	14
7	Loadings for the single-factor model	15
8	Correlation of the four SEIFA indexes for postal areas with SES-C	17
9	SEIFA and SES-C quintiles	20
10	SEIFA + parental occupation and SES-C quintiles	21
11	SEIFA + parental education and SES-C quintiles	21
12	SEIFA + parental occupation + education and SES-C quintiles	21
13	Percentage of quintile participating in higher education by age 19	22
14	Probability of participating in higher education by age 19 by SES measure	22
15	Difference in probabilities between SES values	23
A1	OLS regression of SES-C on SEIFA + parental occupation	26
A2	OLS regression of SES-C on SEIFA + parental education	26
A3	OLS regression of SES-C on SEIFA + parental occupation + education	26
B1	Logistic regression of participating in higher education by age 19 (SES-C)	27
B2	Logistic regression of participating in higher education by age 19 (SEIFA)	27
B3	Logistic regression of participating in higher education by age 19 (SEIFA + parental occupation)	28
B4	Logistic regression of participating in higher education by age 19 (SEIFA + parental education)	28
B5	Logistic regression of participating in higher education by age 19 (SEIFA + parental occupation + education)	29

Figures

1	Scree plot of eigenvalues	15
2	SEIFA and SES-C distribution of differences	19
3	SEIFA-EO and SES-C classification plot	20

Introduction

The measurement of socioeconomic status (SES) is currently receiving considerable policy attention in Australia, particularly in the area of educational participation. This attention is fuelled by the Australian Government's objective to increase access to higher education for youth from low-SES backgrounds (Commonwealth of Australia 2010). Initiatives aimed at low-SES groups are reliant on available measures of SES. Although geographic measures such as the Socio-Economic Indexes for Areas (SEIFA; ABS 2008) are widely used in social and economic research, these measures have been criticised for their imprecision in determining SES at the individual level (Coelli 2010; Jones 2001). Such imprecision can affect both the implementation of policies intended to enhance social participation options for individuals and the monitoring of the effects of these policies.

Our main focus lies on intergenerational mobility. Specifically, we wish to define SES in terms of family characteristics in order to ascertain how the outcomes of young people are affected by their family circumstances. In this context, the Longitudinal Surveys of Australian Youth (LSAY), with its rich set of background characteristics, is an ideal data source to investigate the performance of different measures of SES. If we were interested in social disadvantage for older individuals, then the impact of parental characteristics would be relatively less important compared with other variables in determining SES.

In this paper, we first discuss the different types of information that should be included in an 'ideal' measure of SES for young people. Using LSAY we then create a measure of SES that attempts to approximate this ideal measure, within the constraints of the variables collected in this dataset. Finally, we use our constructed measure of SES as a benchmark against which we assess the performance of several alternative measures of SES. SEIFA is the primary alternative measure for comparison against our derived benchmark. In addition to SEIFA alone, we investigate whether SEIFA can be substantively improved by combining it with other variables (parental occupation and education) which could potentially be collected from student enrolment forms. Alternative measures are assessed by investigating how well they perform in estimating participation in higher education at an aggregate level. We conclude our analysis by examining the effects of the different SES measures in multivariate regression analysis.

We show that use of SEIFA leads to severe misclassification of SES at the individual level. By contrast, SEIFA and SEIFA composites are reasonably accurate measures of SES at the aggregate level, as shown by our examination of aggregate participation in higher education. SEIFA and SEIFA composites also perform acceptably in modelling the relationship between SES and higher education participation, although this relationship is slightly underestimated. Supplementing SEIFA with information on parental occupation or education results in only marginal improvements in individual-level classification. This means that SEIFA and SEIFA composites are inappropriate measures for programs delivered to low-SES individuals, because the majority of such individuals are, in fact, not low-SES.

Relevant dimensions of SES

Socioeconomic status is a multi-dimensional, relative concept that can be measured in a variety of ways. Scutella, Wilkins and Horn (2009) have recently suggested a number of critical dimensions that affect SES and social inclusion. Important dimensions include material resources, social and economic participation, education and health, political or community participation, and access to services. Similar dimensions have been proposed by other researchers (see Pantazis, Gordon & Levitas 2006; Saunders, Naidoo & Griffiths 2007).

When focusing on the SES of young people, it is particularly critical to consider parental background characteristics. These include parental occupation and education, as well as several dimensions of household income and wealth. Individually and collectively, these parental background characteristics determine family access to social, cultural, and economic resources. It is therefore important that these characteristics be reflected by any good measure of individual SES.

The following sections briefly outline ways of measuring some of the important contributors when determining individual-level SES by capturing the impact of such things as parental education and occupation, and household wealth.

Parental occupation

Parental occupation is an important determinant of individual socioeconomic position. The Australian and New Zealand Standard Classification of Occupations (ANZSCO; ABS 2009) represents one prominent classification scheme for occupations in Australia. ANZSCO is a categorical measure that can be used to create a ranking of occupations based on skill level or some related criterion. Alternatively, it is possible to create a continuous measure of occupational prestige by converting ANZSCO classifications to the Australian Socioeconomic Index 2006 (AUSEI06; McMillan, Jones & Beavis 2008). The AUSEI06 scale represents a composite socioeconomic index and reflects the linkages between education, occupation and income.

In addition to domestic occupational classification systems, the International Standard Classification of Occupations (ISCO; International Labour Organization 1990) is frequently used to compare the prestige level of occupations internationally. Occupations coded to ISCO can be converted to the continuous scale of the International Socio-Economic Index of Occupational Status (ISEI; Ganzeboom, De Graaf & Treiman 1992). Continuous occupational index scores such as ISEI are suitable for regression modelling as well as for categorisation into quartiles or quintiles for purposes of cross-tabulation.

Apart from which occupational classification scheme to use, the question of which parent's occupation to measure has also to be decided. Three common options are outlined below:

- ✧ *Focus on father's occupation only*: this approach assumes that the adult male in the household has the strongest attachment to the labour force. However, in the modern labour market this approach probably underestimates family SES, as females nowadays routinely make significant contributions to household income.
- ✧ *Focus on the higher-status occupation*: this approach assumes that the adult with the higher-status occupation determines the family's overall socioeconomic position.

✧ *Focus on father's occupation or, if missing or unknown, the mother's occupation:* this approach is useful in that it helps to overcome missing values that may arise due to male detachment from the labour force. Moreover, it is likely that young people are able to identify the occupation of at least one of their parents or parent figures.

For this reason, we use the third option and map information on parental occupation to the continuous ISEI scale.

Parental education

Parental educational attainment is another important element of an accurate SES measure for young people. In Australia, educational attainment is often classified according to the Australian Qualifications Framework (AQF 2007) or the Australian Standard Classification of Education (ASCED; ABS 2001). An alternative approach to measuring educational attainment is to focus on the length of formal education. However, in the Australian context substantial differences exist in the duration of qualifications, particularly for vocational education and training (VET) courses.

Educational attainment is often classified using the International Standard Classification of Education (ISCED; UNESCO 1997). ISCED facilitates the comparison of education statistics and indicators within and between countries. Table 1 provides an overview of ISCED classifications.

Table 1 ISCED classification

ISCED level	Qualification level
0	Pre-primary, kindergarten, pre-school
1	Primary
2A/B	Certificate I and II (general enabling, bridging courses)
2C	Certificate I and II (basic vocational)
3A/B	Higher school certificate, university enabling courses, AQF certificate III
3C	AQF statement of attainment
4A/B	Certificate IV
5A	Bachelor, bachelor with honours, master (research and coursework)
5B	Diploma, advanced diploma, graduate certificate, graduate diploma
6	PhD, professional doctorate

Similar to the measurement of occupation categories previously discussed, considering parental education raises the question of which parent's educational attainment to measure. Three common options are outlined below:

- 1 *Focus on mother's education only:* this approach is based on an argument of nurture versus nature. Traditionally, mothers provide guidance in child rearing, and a mother who values education is likely to instil this value in her child.
- 2 *Focus on the higher level of education:* this approach assumes that the adult with the higher educational attainment level exerts a leading influence over the family's overall socioeconomic position. Problems arise when the contribution of the adult with the lower level of educational attainment to the family's socioeconomic position is underestimated.
- 3 *Focus on mother's education or, if missing or unknown, the father's education:* this approach helps to alleviate the problem of missing values because it is likely that young people are able to identify the educational attainment level of at least one of their parents or parent figures.

As with parental occupation, we use the third option and map information on parental education to the ISCED scale.

Household income and wealth

Household income and wealth are routinely used as indicators of SES because they represent direct measures of access to economic resources. Surveys such as the Household, Income and Labour Dynamics of Australia (HILDA) and the Census of Population and Housing ask individuals to report their weekly income. However, given that respondents frequently perceive income-related questions as intrusive and that, in any case, adolescents may not know their parents' income or may be unwilling to disclose this information, LSAY refrains from collecting income-related information directly.

As an alternative to direct questions on income and wealth, questions about items that denote wealth through the presence of consumer and cultural items in the household are often used as suitable proxies in survey research (Buchmann 2002). It can be argued that parents who can afford certain household and cultural possessions can also offer their children increased access to social, cultural, and economic resources.

Examples of possession-based and wealth-related measures include the number of rooms in the home, or the presence of literature or art. Factor analytic techniques facilitate the conversion of possession-based responses into a single dimensional measure of wealth. LSAY contains ample information on the presence of consumer and cultural items in the home. We use this information in later sections of this paper.

Ancillary dimensions of SES

Various ancillary dimensions could potentially be included in a measure of SES, including family structure, regionality, immigrant status, and Indigenous status. However, the current literature is divided over the benefits of including any of these ancillary dimensions in a measure of SES. While we discuss each of these ancillary dimensions briefly, our overall conclusion is that they should not be included in a measure of SES.

Family structure

Some evidence suggests that students who live in single-parent or blended families exhibit lower academic achievement when compared with peers from traditional or nuclear families (Marks et al. 2000). This disadvantage may result from differences in family income or time spent with parents or parental figures. Yet, it remains unclear how family structure can be classified within the broader framework of SES, for the simple reason that some single-parent or blended families may have high access to economic and social resources.

Regionality

Regionality denotes whether an individual resides in a metropolitan, regional, or remote area. It is not an area-based measure per se, since it is not concerned with the average economic profile for a given geographic location. Instead, regionality captures the notion that socioeconomic disadvantage can result from the relative distance to resources, such as education providers, libraries, museums, and other infrastructure of educational and cultural importance. Individuals residing in regional or remote areas may be disadvantaged by the distance to such resources, regardless of their personal economic circumstances.

Immigrant status and home language

Strong relationships between SES and immigrant status have been ascertained by researchers in the United States and Europe (Bradley & Corwyn 2002; Crul & Vermeulen 2003). However, such relationships have not been determined for the Australian context (Cresswell 2004). The benefits of including immigrant status in a measure of SES thus remain unclear.

Indigenous status

Although Indigenous status is associated with low SES (Centre for the Study of Higher Education 2008), the use of this variable as a measure of SES is debatable. Ambiguity arises from the fact that an Indigenous individual may have a relatively strong economic position, yet still be at a distinct disadvantage relative to other groups in society when non-economic factors are considered. The benefits of including Indigenous status in a measure of SES thus remain unclear.

Towards creating a measure of SES

LSAY contains a rich set of variables that represent some of the dimensions that Scutella, Wilkins and Horn (2009) have identified. These dimensions include parental education and occupation, as well as measures of possessions in the family home. In the following section, we use the most important SES-related variables to create a reference measure of SES from the data that are currently available in LSAY.

Creating a reference measure of SES

We use data from the LSAY 2003 cohort to create a reference measure for SES. LSAY is a nationally representative survey that tracks young people from the age of 15 to 25 years as they move from school into further study, work, and other destinations. LSAY contains a rich set of individual background variables, which is an important prerequisite for creating an accurate SES reference measure. We use a set of 16 SES-related background variables as a basis for creating our SES reference measure (see table 2).

Table 2 LSAY variables used as a basis for creating an SES reference measure

Variable	Type or categories	Valid n	\hat{p}	SE
Own desk at home	Dichotomous	10 366	0.902	0.003
Own room at home	Dichotomous	10 366	0.915	0.003
Own study place at home	Dichotomous	10 367	0.834	0.004
Computer software at home	Dichotomous	10 366	0.669	0.005
Internet at home	Dichotomous	10 366	0.849	0.003
Calculator at home	Dichotomous	10 366	0.971	0.002
Literature at home	Dichotomous	10 365	0.363	0.005
Poetry at home	Dichotomous	10 366	0.406	0.005
Art at home	Dichotomous	10 366	0.556	0.005
Textbooks at home	Dichotomous	10 366	0.802	0.004
Dictionary at home	Dichotomous	10 366	0.973	0.002
Dishwasher at home	Dichotomous	10 362	0.594	0.005
Number of books at home	0 – 10	447	0.043	0.002
	11 – 25	883	0.085	0.003
	26 – 100	2 848	0.275	0.004
	101 – 200	2 347	0.226	0.004
	201 – 500	2 205	0.213	0.004
	More than 500	1 486	0.143	0.003
Parental occupation	Missing	107	0.015	0.001
	Continuous	9 417	46.635*	17.246
Parental education	None	310	0.031	0.002
	ISCED 1	84	0.01	0.001
	ISCED 2	2 064	0.206	0.004
	ISCED 3B,C	319	0.032	0.002
	ISCED 3A, 4	3 174	0.317	0.005
	ISCED 5B	1 201	0.119	0.003
	ISCED5A, 6	2 872	0.286	0.005
Own computer at home	Dichotomous	10 364	0.939	0.002

Notes: Sample sizes and proportions are unweighted. Proportions (\hat{p}) represent the per cent respondents to whom the listed variable condition applies.

*Value represents the mean, not the proportion.

We take a factor-analytic approach to derive our SES reference measure. In an initial step, we factor-analyse the 16 SES-related variables listed in table 2 into a smaller number of distinct factors. Traditional factor analysis assumes all of the variables included in the model to be continuous. It

further assumes that the emerging factors of interest follow a continuous distribution. In our case, however, the majority of the 16 variables used in the model are dichotomous or categorical, thereby violating basic distributional assumptions for traditional factor analysis. We remedy this issue by using tetrachoric (dichotomous) and polychoric (categorical) correlations in a latent class factor analysis, which is a better fit for the particular distributional properties of our SES-related variables. Eigenvalues and related statistics are provided in table 3. The interpretation of results from latent-class factor analysis is similar to that of traditional factor analysis.

Table 3 Initial latent-class factor analysis

Factor	Eigenvalue	Difference	Prop. explained	Cumulative
1	6.28	NA	39.25	39.25
2	1.74	4.54	10.89	50.14
3	1.48	0.26	9.26	59.39
4	1.15	0.34	7.16	66.55
5	0.81	0.33	5.08	71.63
6	0.72	0.09	4.52	76.15
7	0.71	0.02	4.41	80.56
8	0.65	0.06	4.06	84.61
9	0.54	0.11	3.37	87.98
10	0.48	0.06	2.97	90.95
11	0.42	0.06	2.61	93.56
12	0.35	0.07	2.18	95.74
13	0.29	0.06	1.81	97.55
14	0.21	0.08	1.32	98.87
15	0.17	0.04	1.07	99.94
16	0.01	0.16	0.06	100.00

Note: NA = not applicable

We observe four factors with eigenvalues in excess of 1 (in bold). Collectively, these four factors explain over 66% of the total variance in the model. We call the four factors income and wealth, study resources, computing resources, and cultural resources (see table 4).

Table 4 Loadings for the four-factor model (varimax rotation)

LSAY variable	Study resources	Income/wealth	Computing resources	Cultural resources
Desk	0.628	0.167	0.243	0.259
Own room	0.524	0.312	-0.026	-0.052
Study place	0.758	0.100	0.097	0.257
Software	0.413	-0.043	0.547	0.377
Internet	0.251	0.390	0.641	0.062
Calculator	0.675	-0.033	0.321	0.206
Literature	0.148	0.319	0.136	0.787
Poetry	0.134	0.172	0.040	0.880
Art	0.231	0.149	0.116	0.619
Textbooks	0.384	0.023	0.214	0.559
Dictionary	0.667	0.012	0.353	0.382
Dishwasher	0.237	0.422	0.272	0.089
Parental occupation	0.060	0.507	0.101	0.165
No. of books	0.109	0.305	0.098	0.453
Parental education	0.005	0.503	0.068	0.227
Computer in home	0.198	0.279	1.109	0.171

Note: The highest loading for each variable across all four factors is in bold.

From the data in table 4 we observe that the variable capturing computer availability in the home produces a loading in excess of 1, indicating an estimation anomaly. A likely cause for this anomaly is collinearity between the presence of a computer in the home, the availability of software, and internet access. To obviate further estimation problems, we eliminate the ‘computer in the home’ variable from the model. Results from the latent-class factor analysis for the modified model are provided in table 5.

Table 5 Modified latent-class factor analysis model

Factor	Eigenvalue	Difference	Proportion of variance explained	Cumulative
1	5.75	NA	38.32	38.32
2	1.68	4.07	11.20	49.52
3	1.38	0.30	9.20	58.72
4	0.94	0.44	6.25	64.97
5	0.78	0.15	5.23	70.19
6	0.71	0.08	4.72	74.91
7	0.70	0.01	4.69	79.60
8	0.62	0.08	4.15	83.75
9	0.53	0.09	3.53	87.28
10	0.48	0.06	3.17	90.45
11	0.42	0.06	2.78	93.23
12	0.35	0.07	2.31	95.54
13	0.29	0.06	1.93	97.47
14	0.21	0.08	1.41	98.83
15	0.17	0.05	1.12	100.00

The modified model identifies three factors with eigenvalues greater than 1 (in bold). The three-factor model in table 6 identifies educational resources, income and wealth, and cultural resources as underlying traits of SES.

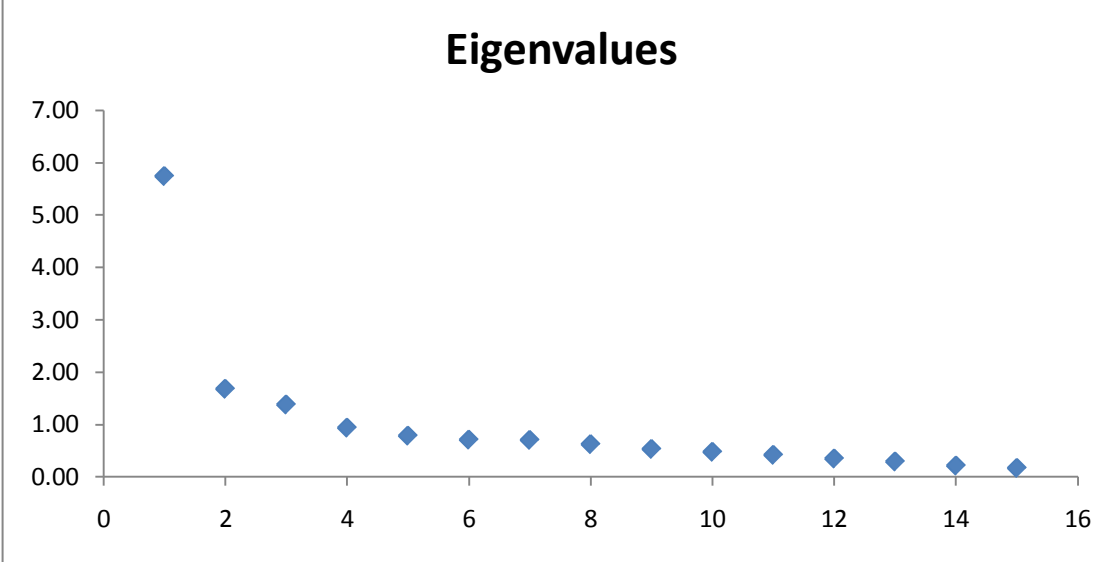
Table 6 Loadings for the three-factor model

LSAY variable	Educational resources	Income/wealth	Cultural resources
Desk	0.664	0.219	0.229
Own room	0.396	0.291	-0.032
Study place	0.685	0.130	0.231
Software	0.608	0.091	0.324
Internet	0.464	0.493	0.063
Calculator	0.776	0.052	0.141
Literature	0.210	0.302	0.791
Poetry	0.176	0.128	0.883
Art	0.287	0.146	0.608
Textbooks	0.484	0.057	0.524
Dictionary	0.788	0.091	0.323
Dishwasher	0.299	0.482	0.083
Parental occupation	0.047	0.515	0.188
No. of books	0.132	0.293	0.463
Parental education	-0.014	0.495	0.256

Note: The highest loading for each variable across all three factors is in bold.

Although three factors with eigenvalues greater than 1 are identified, attention should be given to the sizeable difference between the eigenvalue of the first factor and those of the other two factors. Figure 1 underscores the magnitude of this difference.

Figure 1 Scree plot of eigenvalues



Given the disproportionately large amount of variance explained by the first factor, we isolate this factor to generate one composite measure of SES consisting of a variety of home resources and parental background dimensions (see table 7).

Table 7 Loadings for the single-factor model

LSAY Variable	Composite SES Factor
Desk	0.628
Own room	0.340
Study place	0.618
Software	0.598
Internet	0.499
Calculator	0.640
Literature	0.844
Poetry	0.803
Art	0.653
Textbooks	0.659
Dictionary	0.782
Dishwasher	0.426
No. of books	0.523
Parental occupation	0.359
Parental education	0.380

We use the single-factor model as our SES reference measure for the remainder of this investigation. Our decision is based on practical considerations related to the interpretability and useability of subsequent analyses. Given that our primary interest centres on the evaluation of

various different measures of SES, the composite SES measure from the single-factor model provides a less complex reference for comparison. For the remainder of this paper, we refer to our SES reference measure as SES-C (SES-Composite).¹

¹ Our SES measure overlaps to some extent with the Index of Economic, Social, and Cultural Status (ESCS), a measure of SES developed by the Organisation for Economic Co-operation and Development (OECD 2005) for use in the Program for International Student Assessment (PISA). Similar to our measure, ESCS is derived from family background variables, including parental occupation, parental education, and home possessions. Despite the high correlation between our SES-C measure and ESCS ($r = 0.75$), important differences exist. ESCS scores are obtained as component scores for the first principal component from factor analysis, whereby θ is the score of an average OECD student and σ the standard deviation across equally weighted OECD countries. The need for multi-country adjustment renders ESCS less reliable when considering only the Australian context. This loss in reliability is reflected in the considerably lower reliability coefficient for ESCS compared with our SES-C reference measure (standardised Cronbach's alpha for ESCS for Australia = 0.61; standardised Cronbach's alpha for SES-C = 0.74).

Considering alternative measures of SES

Socio-Economic Indexes for Areas

Area-based measures are often used as proxies for SES. In Australia, the Socio-Economic Indexes for Areas (SEIFA; ABS 2008) are among the most prominent area-based measures of SES. SEIFA use information from the five-yearly Census of Population and Housing to determine the presence of socioeconomic advantage or disadvantage. SEIFA indexes are *relative* measures, meaning that a given geographic area may only be determined as disadvantaged *relative to other areas* rather than disadvantaged *in absolute terms*. SEIFA comprises four separate indexes, each of which represents a slightly different approach to measuring SES.

- 1 *Index of Relative Socioeconomic Disadvantage (IRSD)*: this index summarises 17 area-based variables, including low income, low education, as well as high rates of unemployment and unskilled occupations.
- 2 *Index of Relative Socioeconomic Advantage and Disadvantage (IRSAD)*: this index features 21 area-based variables, including low or high income, internet connection, occupation and education.
- 3 *Index of Economic Resources (IER)*: this index contains 15 area-based variables, including items such as household income, housing expenditure, and wealth.
- 4 *Index of Education and Occupation (IEO)*: this index is composed of nine area-based variables, including educational attainment, enrolment in further education, occupational information (such as skill level), and unemployment status.

Each of the four SEIFA indexes is available for a range of different geographical entities, such as collection districts, statistical local areas, local government areas, state suburbs, and postal areas. We exclusively consider indexes for postal areas because information contained in the LSAY dataset is limited to respondents' residential postcode. Correlations between our SES-C reference measure and each of the four indexes for postal areas are weak but highly similar (see table 8).

Table 8 Correlation of the four SEIFA indexes for postal areas with SES-C

Variable	Correlation with SES-C
SEIFA Index of Education and Occupation (for postal areas)	0.30
SEIFA Index of Economic Resources (for postal areas)	0.26
SEIFA Index of Relative Advantage (for postal areas)	0.29
SEIFA Index of Relative Disadvantage (for postal areas)	0.29

For the remainder of our analysis we use the most highly correlated SEIFA Index of Education and Occupation for postal areas as a basis for comparison. This index is henceforth referred to as simply SEIFA.

SEIFA composites

When determining the SES of young people, measures that consider key background characteristics of parents or parent figures can increase classification accuracy. Information on parental occupation or education offers two particular advantages. First, questions about parental occupation and/or education can be added to student enrolment forms with relative ease. Second, such questions are generally perceived by respondents as less intrusive than direct income-related questions.

We create three SEIFA composites as a basis for testing whether supplementary information on parental occupational or educational background can enhance classification performance in area-based measures of SES. The three SEIFA composites include:

- ✧ SEIFA + parental occupation (using the International Socio-Economic Index of Occupational Status)
- ✧ SEIFA + parental education (using the International Standard Classification of Education)
- ✧ SEIFA + parental occupation + education.

To create these measures, we conduct a series of regressions, whereby SES-C is regressed against the factors of each SEIFA composite. Predicted mean scores are computed and used as the respective SEIFA composite measure (appendix A).

Testing alternative SES measures

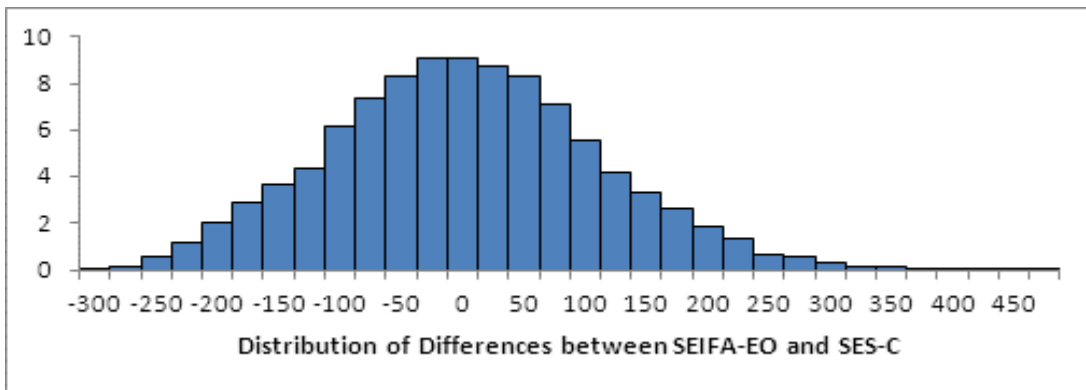
In this section, we test the classification performance of alternative SES measures by comparing them to our SES-C reference measure. Alternative SES measures include SEIFA, SEIFA + parental occupation, SEIFA + parental education, and SEIFA + parental occupation and education. To carry out meaningful comparisons, SES-C is assumed to be a reference measure that correctly classifies individuals into their respective SES categories. Alternative SES measures are tested at the individual and aggregate levels, and as predictors in multivariate analysis.

Performance of SES measures at the individual level

SEIFA

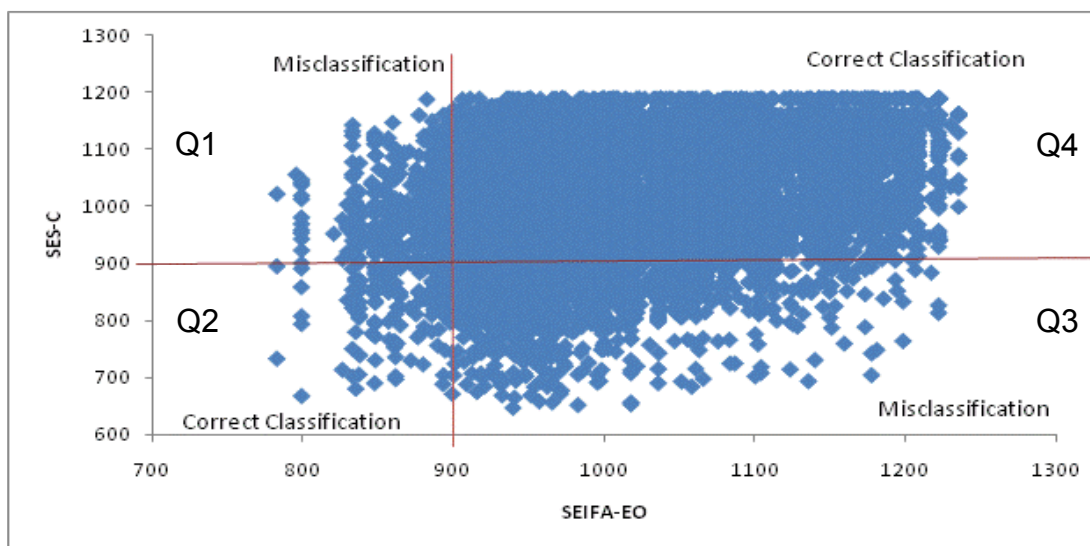
We use three different approaches to assess the classification performance of SEIFA against that of SES-C. Our first approach consists of examining the distribution of the differences between the two measures, whereby factor scores from the one-factor model are standardised ($m = 1000$; $sd = 100$). This step allows us to use standard deviation units for comparison by imposing identical distributions on SEIFA and SES-C scores. The distribution of differences between the scores is provided in figure 2. While the majority of the differences fall between -1.5 and +1.5 standard deviations, substantial differences are observed well beyond this range.

Figure 2 SEIFA and SES-C distribution of differences



As a second approach, we create a classification plot for comparing SEIFA against SES-C (figure 3). The low correlation between the two measures is clearly reflected in the shape of the plot. Assuming a score of 900 or below (that is, one standard deviation or more below the SES mean score) to represent low SES, we examine the classification of low-SES individuals by each of the two measures. Using SES-C classifications as the reference, quadrants 2 and 4 depict the substantial number of misclassified individuals when using SEIFA. However, in the main, these quadrants show individuals who are correctly classified using both SES-C and SEIFA. Quadrant 1 shows actual high-SES individuals who are misclassified as low SES when using SEIFA, but are in fact not low SES when using SES-C, and quadrant 3 shows those incorrectly classified as being from high SES when using SEIFA, but are low SES when classified using SES-C.

Figure 3 SEIFA-EO and SES-C classification plot



Cross-tabulation of quintiles is a third approach with which we assess the classification accuracy of SEIFA when compared with the SES-C reference measure. In the cross-tabulation matrix (table 9), the sum of values along the diagonal vector represents the per cent correct classification rate. Quintile values of 20 along the diagonal vector would sum to a 100% correct classification rate and indicate a perfect relationship between SEIFA and SES-C. The actual matrix illustrates a high level of misclassification, whereby the diagonal sum indicates a correct classification rate of only 26%. An examination of first-order off-diagonal vectors, which denote slight deviations from correct classification, indicates that 35% of individuals are somewhat misclassified by SEIFA. Finally, second-order and above off-diagonal vectors, which denote substantial deviations from correct classification, show that SEIFA severely misclassifies almost 40% of the sample. Our misclassification results confirm findings by Coelli (2010), who compared SEIFA’s SES classifications with data on income levels using the Household, Income and Labour Dynamics in Australia Survey. Table 9 presents the cross-tabulation for SEIFA and SES-C quintiles.

Table 9 SEIFA and SES-C quintiles

SES-C	SEIFA					Total
	1	2	3	4	5	
1 = lowest	5.53	5.19	4.09	3.38	1.80	20
2	4.57	4.51	4.34	4.04	2.52	20
3	4.20	4.24	4.29	4.03	3.30	20
4	3.54	3.72	4.04	4.03	4.65	20
5 = highest	2.03	2.42	3.59	4.30	7.64	20
Total	20.00	20.00	20.00	20.00	20.00	100

Given the clarity and ease of interpretability of cross-tabulations, we focus on this approach to assess the classification performance of the remaining alternative SES measures in the following sections.

SEIFA + parental occupation

An examination of the cross-tabulation matrix for SEIFA + parental occupation and SES-C indicates a correct classification rate of 28.5% (sum of diagonal vector) and a slight misclassification rate of 37% (first-order off-diagonals). However, 34.5% of the sample remain severely misclassified (second-order and above off-diagonals). While adding information on parental occupation yields an improvement in classification accuracy, overall misclassification remains strong. Table 10 presents the cross-tabulation for SEIFA + parental occupation and SES-C quintiles.

Table 10 SEIFA + parental occupation and SES-C quintiles

SES-C	SEIFA + parental occupation					Total
	1	2	3	4	5	
1	6.24	5.21	3.87	2.47	1.28	20
2	4.90	4.78	4.49	3.43	2.44	20
3	4.10	4.38	4.35	4.31	3.11	20
4	3.35	3.48	3.85	4.73	4.85	20
5	1.40	2.15	3.42	5.08	8.33	20
Total	20.00	20.00	20.00	20.00	20.00	100

SEIFA + parental education

When supplementing SEIFA with information on parental education, the correct classification rate increases marginally to 30.7%. Slight and severe misclassifications are at 37% and 32.3%, respectively. It is interesting that adding information on parental education results in a stronger improvement of classification accuracy when compared with adding information on parental occupation. Nonetheless, overall misclassification remains strong despite slight improvements. Table 11 presents the cross-tabulation for SEIFA + parental education and SES-C quintiles.

Table 11 SEIFA + parental education and SES-C quintiles

SES-C	SEIFA + parental education					Total
	1	2	3	4	5	
1	7.31	5.05	3.91	2.30	1.04	20
2	5.27	4.95	4.43	3.13	2.12	20
3	3.47	4.41	4.46	4.40	3.35	20
4	2.90	3.75	3.78	5.13	4.66	20
5	1.09	1.84	3.37	5.04	8.85	20
Total	20.00	20.00	20.00	20.00	20.00	100

SEIFA + parental occupation + parental education

The combination of SEIFA with information on parental occupation and education yields a correct classification rate of 31.3%, representing only a marginal improvement compared with the previous education-only measure. Slight misclassification increased to 38%, and severe misclassification decreased to 30% (table 12).

Table 12 SEIFA + parental occupation + education and SES-C quintiles

SES-C	SEIFA-EO + parental occupation + education					Total
	1	2	3	4	5	
1	7.37	5.19	3.29	2.14	0.82	20
2	5.43	5.01	4.38	3.15	1.93	20
3	3.46	4.50	4.67	4.65	3.00	20
4	2.89	3.40	4.17	5.00	4.96	20
5	0.85	1.88	3.49	5.05	9.29	20
Total	20.00	20.00	20.00	20.00	20.00	100

Performance of SES measures at the aggregate level

The above cross-tabulations have illustrated the strength of classification bias when SEIFA and SEIFA composites are used for SES categorisations at the individual level. The cross-tabulations presented in the previous section do not provide insight into the performance of these measures at

an aggregate level. To assess aggregate-level performance, we consider the SES classifications of each measure with regards to a particular outcome of interest. Specifically, we examine higher education participation rates by age 19 years. Based on our LSAY sample, table 13 presents the higher education participation rates by quintile for SES-C and each of the alternative measures.

Table 13 Percentage of quintile participating in higher education by age 19

Quintile	SES-C	SEIFA	SEIFA + par. occ.	SEIFA + par. ed.	SEIFA + par. occ. + ed.
1	20.07	26.53	25.57	22.01	21.51
2	30.50	30.07	26.58	29.36	28.09
3	37.46	33.96	33.32	32.39	33.80
4	42.59	40.31	45.16	46.42	45.45
5	55.00	60.31	61.81	60.85	63.09

While SEIFA and SEIFA composites overstate participation rates in the lowest and highest SES quintiles, deviations from SES-C remain within relatively moderate bounds. Overall, table 13 demonstrates that SEIFA and SEIFA composites produce satisfactory results at the aggregate classification level, despite performing poorly at the individual classification level.

Performance of SES measures in multivariate analysis

We conclude the analysis by determining how well the different SES measures perform in multivariate models. A logistic regression model for each of the different SES measures is fitted to predict the probability of participation in higher education by age 19 years. Each regression model contains a set of additional variables known to influence participation. These variables include gender, Indigenous status, metropolitan region status, secondary school sector, family structure, as well as academic achievement in mathematics, reading, and science.

Results from the regression analyses are presented as predicted probabilities of participating in higher education (see table 14). Predicted probabilities are determined for individuals classified as low, medium, and high-SES. Individuals are classified into the three SES categories based on SES-C quintiles (detailed regression results are provided in appendix B).²

Table 14 Probability of participating in higher education by age 19 by SES measure

SES measure	Low SES predicted probability	Medium SES predicted probability	High SES predicted probability
SES-C	0.230	0.374	0.570
SEIFA	0.267	0.383	0.542
SEIFA + parental occpn	0.267	0.384	0.547
SEIFA + parental educn	0.250	0.386	0.560
SEIFA + parental occpn + educn	0.254	0.387	0.563

Table 14 demonstrates that, when compared with our SES-C reference measure, all alternative measures are biased in terms of predicting higher education participation. SEIFA and SEIFA composites overstate participation probabilities for individuals in low and medium-SES categories,

² Probabilities are calculated for remaining variables at their reference levels and at the average for continuous variables (mathematics, science and reading achievement). Low, middle and high SES are derived at the values of the first, third and fifth quintile.

while marginally understating probabilities for high-SES individuals. We further note that SEIFA + parental occupation + education performs well for predicting participation for high-SES individuals.

Table 15 presents the differences in probabilities between high and low, medium and low, and high and medium SES individuals. It is apparent that using the SEIFA measures in regression will understate the strength of the relationship between SES and higher education participation. Therefore, if a relationship was ascertained statistically, we could be fairly confident that it actually exists (note that standard errors have not been compared).

Table 15 Difference in probabilities between SES values

SES measure	High–low	High–medium	Medium–low
SES-C	0.340	0.197	0.144
SEIFA	0.276	0.160	0.116
SEIFA + parental occpn	0.281	0.163	0.118
SEIFA + parental educn	0.310	0.174	0.136
SEIFA + parental occpn + educn	0.309	0.175	0.134

Overall, alternative SES measures perform reasonably well in multivariate regressions for individuals in the medium or high end of the SES distribution.

Conclusion

For young people, an accurate SES measure should capture a rich set of parental background characteristics that mirror access to social, cultural, and economic resources. The limited availability of such information frequently necessitates the use of area-based SEIFA indexes as proxies for SES. Against this backdrop, we have created a reference measure that contains the rich set of parental background variables necessary to accurately classify youth into SES categories. Subsequently, we have considered available alternative SES measures and tested the performance of alternative SES measures at individual and aggregate levels.

Our analysis has produced the following key findings:

- ✧ SEIFA greatly misclassifies SES at the individual level.
- ✧ Supplementing SEIFA with information on parental occupation or education results in only marginal improvements of individual-level classification accuracy.
- ✧ SEIFA and SEIFA composites perform satisfactorily with respect to reporting participation in higher education at aggregate levels.
- ✧ SEIFA and SEIFA composites work reasonably in modelling the relationship between SES and higher education participation, although this relationship is slightly underestimated.

Our overall conclusion is that SEIFA is satisfactory for aggregate relationships but results in serious misclassification at the individual level. The implication is that it is an inappropriate measure for programs delivered to low-SES individuals, because the majority of such individuals are, in fact, not low-SES. Moreover, supplementing SEIFA with data on parental occupation or education does not lead to substantive improvements of individual classification accuracy.

References

- ABS (Australian Bureau of Statistics) 2001, *Australian standard classification of education (ASCED)*, Commonwealth of Australia, Canberra.
- 2008, *An introduction to socio-economic indexes for areas (SEIFA)*, Commonwealth of Australia, ABS, Canberra.
- 2009, *ANZSCO—Australian and New Zealand Standard Classifications of Occupations*, ABS, Canberra.
- AQF (Australian Qualifications Framework) 2007, *Implementation Handbook 4th Edition*, MCEETYA, Melbourne.
- Bradley, RH & Corwyn, RF 2002, 'Socioeconomic status and child development', *Annual Review of Psychology*, vol.53, pp.371–99.
- Buchmann, C 2002, 'Measuring family background in international studies of education: conceptual issues and methodological challenges', in *Methodological advances in cross-national surveys of educational achievement*, eds AC Porter & A Gamoran, National Academy Press, Washington, pp.150–97.
- Centre for the Study of Higher Education (CSHE) 2008, *Participation and equity: a review of the participation in higher education of people from low socioeconomic backgrounds and Indigenous people*, University of Melbourne, Melbourne.
- Coelli, MB 2010, *The forgotten second quartile: parental income and youth post-secondary education enrolment in Australia*, University of Melbourne, Melbourne.
- Commonwealth of Australia 2010, *Higher education participation and partnerships program*, viewed 9 March 2010, <http://www.deewr.gov.au/HigherEducation/Programs/Equity/Documents/HEPPPGuidelines_2010.pdf>.
- Cresswell, J 2004, *Immigrant status and home language background: implications for Australian student performance in PISA 2000*, Australian Council for Educational Research, Camberwell.
- Crul, M & Vermeulen, H 2003, 'The second generation in Europe', *International Migration Review*, vol.37, pp.965–86.
- Ganzeboom, HBG, De Graaf, PM & Treiman, DJ 1992, 'A standard international socio-economic index of occupational status', *Social Science Research*, vol.21, issue 1, pp.1–56.
- International Labour Organization 1990, *International standard classifications of occupations (ISCO-88)*, ILO, Geneva, viewed January 2011, <<http://www.ilo.org/public/english/bureau/stat/isco/index.htm>>.
- Jones, RG 2001, *Identifying higher education students from low socio-economic status backgrounds and regional and remote areas*, Department of Education, Science and Training, Canberra.
- Marks, GN, McMillan, J, Jones, FL & Ainley, J 2000, *The measurement of socioeconomic status for the reporting of nationally comparable outcomes of schooling*, Australian Council for Educational Research, Camberwell.
- McMillan, J, Jones, F & Beavis, A 2008, *Introducing the Australian socioeconomic index 2006*, Australian Demographic and Social Research Institute, Canberra.
- OECD (Organisation for Economic Co-operation and Development) 2005, *PISA 2003 Technical report*, OECD, Paris, pp.316–19.
- Pantazis, C, Gordon, D & Levitas, R (eds) 2006, *Poverty and social exclusion in Britain: the Millennium survey*, Policy Press, Bristol.
- Saunders, P, Naidoo, Y & Griffiths, M 2007, *Towards new indicators of disadvantage: deprivation and social exclusion in Australia*, Social Policy Research Centre, Sydney.
- Scutella, R, Wilkins, R & Horn, M 2009, *Measuring poverty and social exclusion in Australia: a proposed multidimensional framework for identifying socio-economic disadvantage*, University of Melbourne, Melbourne.
- UNESCO (United Nations Educational, Scientific and Cultural Organization) 1997, *International Standard Classification of Education (ISCED)*, Paris.

Appendix A

Table A1 OLS regression of SES-C on SEIFA + parental occupation

Parameter		Estimate	SE	t	Df
R ²	0.142				
Intercept		693.96	10.85	63.94***	1
SEIFA		0.24	0.01	21.26***	1
Parental occpn		1.41	0.06	24.63***	1

Note: * p < 0.05 ** p < 0.01 *** p < 0.001

Table A2 OLS regression of SES-C on SEIFA + parental education

Parameter		Estimate	SE	t	Df
R ²	0.193				
Intercept		816.52	11.45	71.32***	1
SEIFA		0.23	0.01	21.14***	1
ISCED – none		-91.83	5.39	-16.84***	1
ISCED 1		-132.25	9.83	-13.45***	1
ISCED 2		-86.30	2.69	-32.15***	1
ISCED 3B, C		-70.47	5.30	-13.30***	1
ISCED 3A, 4		-54.99	2.37	-23.20***	1
ISCED 5B		-36.49	3.13	-11.67***	1
ISCED 5A, 6		Reference level			

Note: * p < 0.05 ** p < 0.01 *** p < 0.001

Table A3 OLS regression of SES-C on SEIFA + parental occupation + education

Parameter		Estimate	SE	t	Df
R ²	0.218				
Intercept		827.22	11.35	72.90***	1
SEIFA		0.17	0.01	14.86***	1
Parental occpn		1.03	0.06	18.33***	1
ISCED – none		-76.96	5.51	-13.98***	1
ISCED 1		-117.44	10.27	-11.44***	1
ISCED 2		-75.70	2.71	-27.94***	1
ISCED 3B, C		-60.51	5.30	-11.41***	1
ISCED 3A, 4		-47.92	2.37	-20.23***	1
ISCED 5B		-30.11	3.10	-9.73***	1
ISCED 5A, 6		Reference level			

Note: * p < 0.05 ** p < 0.01 *** p < 0.001

Appendix B

Table B1 Logistic regression of participating in higher education by age 19 (SES-C)

Parameter	Estimate	SE	Wald χ^2	Df
<i>Per cent concordant</i>	79			
Intercept	-8.523	0.692	151.69***	1
SES-C	0.002	0.000	28.36***	1
Male	-0.630	0.075	70.98***	1
Non-Indigenous	0.389	0.193	4.04*	1
Metropolitan	-0.372	0.077	23.47***	1
Government	-0.747	0.095	62.43***	1
Catholic	-0.309	0.107	8.33**	1
Single family	-0.286	0.504	0.32	1
Nuclear family	0.163	0.499	0.11	1
Mixed family	-0.623	0.517	1.45	1
Other family	-0.010	0.558	0.00	1
Maths scores	0.007	0.001	93.36***	1
Reading scores	0.004	0.001	23.32***	1
Science scores	0.001	0.001	2.34	1

Note: * p < 0.05 ** p < 0.01 *** p < 0.001

Table B2 Logistic regression of participating in higher education by age 19 (SEIFA)

Parameter	Estimate	SE	Wald χ^2	Df
<i>Per cent concordant</i>	78.9			
Intercept	-14.78	1.438	105.68***	1
SEIFA EO	0.008	0.001	37.16***	1
Male	-0.640	0.075	72.22***	1
Non-Indigenous	0.394	0.196	4.06*	1
Metropolitan	-0.205	0.083	6.19*	1
Government	-0.688	0.098	49.45***	1
Catholic	-0.269	0.110	5.99*	1
Single family	-0.333	0.496	0.45	1
Nuclear family	0.180	0.491	0.13	1
Mixed family	-0.638	0.510	1.57	1
Other family	-0.030	0.548	0.00	1
Maths scores	0.007	0.001	88.60***	1
Reading scores	0.004	0.001	22.30***	1
Science scores	0.002	0.001	3.43	1

Note: * p < 0.05 ** p < 0.01 *** p < 0.001

Table B3 Logistic regression of participating in higher education by age 19 (SEIFA + parental occupation)

Parameter	Estimate	SE	Wald χ^2	Df
<i>Per cent concordant</i>	79			
Intercept	-13.705	1.186	133.47***	1
SEIFA + parental occpn	0.007	0.001	43.36***	1
Male	-0.644	0.076	71.15***	1
Non-Indigenous	0.317	0.196	2.62	1
Metropolitan	-0.217	0.082	6.95**	1
Government	-0.680	0.098	47.79***	1
Catholic	-0.241	0.111	4.71*	1
Single family	-0.211	0.469	0.20	1
Nuclear family	0.284	0.464	0.38	1
Mixed family	-0.573	0.484	1.40	1
Other family	-0.047	0.524	0.01	1
Maths scores	0.007	0.001	87.51***	1
Reading scores	0.004	0.001	22.46***	1
Science scores	0.001	0.001	2.59	1

Note: * p < 0.05 ** p < 0.01 *** p < 0.001

Table B4 Logistic regression of participating in higher education by age 19 (SEIFA + parental education)

Parameter	Estimate	SE	Wald χ^2	Df
<i>Per cent concordant</i>	79.4			
Intercept	-14.596	1.040	196.88***	1
SEIFA + parental educn	0.008	0.001	85.92***	1
Male	-0.690	0.075	83.48***	1
Non-Indigenous	0.447	0.201	4.97*	1
Metropolitan	-0.220	0.080	7.61**	1
Government	-0.638	0.098	42.77***	1
Catholic	-0.220	0.110	4.00*	1
Single family	-0.280	0.528	0.28	1
Nuclear family	0.248	0.523	0.23	1
Mixed family	-0.574	0.541	1.13	1
Other family	0.051	0.582	0.01	1
Maths scores	0.007	0.001	89.72***	1
Reading scores	0.003	0.001	18.57***	1
Science scores	0.001	0.001	2.53	1

Note: * p < 0.05 ** p < 0.01 *** p < 0.001

Table B5 Logistic regression of participating in higher education by age 19 (SEIFA + parental occupation + education)

Parameter	Estimate	SE	Wald χ^2	Df
<i>Per cent concordant</i>	79			
Intercept	-14.224	1.011	197.77***	1
SEIFA + parental educn + occpn	0.008	0.001	80.86***	1
Male	-0.687	0.076	81.00***	1
Non-Indigenous	0.372	0.200	3.46	1
Metropolitan	-0.218	0.081	7.22**	1
Government	-0.633	0.099	41.13***	1
Catholic	-0.198	0.112	3.15	1
Single family	-0.173	0.500	0.12	1
Nuclear family	0.349	0.495	0.50	1
Mixed family	-0.480	0.515	0.87	1
Other family	-0.096	0.556	0.03	1
Maths scores	0.007	0.001	85.56***	1
Reading scores	0.003	0.001	18.86***	1
Science scores	0.001	0.001	2.60	1

Note: * p < 0.05 ** p < 0.01 *** p < 0.001

National Centre for Vocational Education Research Ltd
Level 11, 33 King William Street, Adelaide, South Australia
PO Box 8288, Station Arcade, SA 5000 Australia

Telephone +61 8 8230 8400 Facsimile +61 8 8212 3436
Website www.ncver.edu.au Email ncver@ncver.edu.au

