

**Report on the generation of the 2012 *Index of
Community Socio-educational Advantage (ICSEA)***

March 2013

© Australian Curriculum, Assessment and Reporting Authority 2013

This work is copyright. You may download, display, print and reproduce this material in unaltered form only (retaining this notice) for your personal, non-commercial use or use within your organisation.

All images are indicative only and may be subject to change. Graphical depictions should not be relied upon for accuracy of content.

All other rights are reserved. Requests and inquiries concerning reproduction and rights should be addressed to:

ACARA Copyright Administration, ACARA
Level 10, 255 Pitt Street
Sydney NSW 2000

Contents

ICSEA Overview	1
Stage 1: Constructing the school performance scale	3
Stage 2: Constructing the SEA measures	4
2.1 Direct SEA scale	4
2.2 Student enrolment records	4
2.3 Determining the composition of the direct parent data variable set	5
2.4 Simplifying the direct parent data variable set	5
2.5 Indirect SEA scale	6
2.6 Aligning direct and indirect SEA scales	8
Stage 3: Constructing the ICSEA	9
3.1 Inclusion of school variables	9
3.2 Determining the source of the reported ICSEA	9
3.3 Data sufficiency rules	9
3.4 Comparison of distributions of 2012 and 2011 ICSEA	10
Appendix	11
List of modelled 2011 SEIFA Variables	11
Direct parent data SEA equation	12
Indirect parent data SEA equation	12

Illustrations

Table 1. Analyses for SEA calculated using direct data	6
Table 2. Model selection and proposed variables for the calculation of indirect ICSEA	7
Table 3. Scaling factors used to re-scale indirect scale values between adjacent decile medians	8
Table 4. Rules and functions to calculate minimum sufficient number of school records	9
Figure 1. Distribution of 2011 and 2012 ICSEA values	10
Table 5. Modelled 2011 SEIFA Variables	11

ICSEA Overview

The *Index of Community Socio-educational Advantage* (ICSEA) is a scale that represents levels of educational advantage. A value on the scale assigned to a school is the average level for all students in the particular school. ICSEA is calculated by the Australian Curriculum, Assessment and Reporting Authority (ACARA).

ICSEA was developed to enable fair and meaningful comparisons of the literacy and numeracy performance of students in a given school, as assessed by the National Assessment Program—Literacy and Numeracy (NAPLAN) tests, with the performance of students at schools catering for statistically similar students in terms of the level of socio-educational advantage (SEA).

The development of ICSEA involves collecting student family background data and identifying, through the use of regression models, the combination of variables that have the strongest association with student performance, and within that combined grouping, how much each of those variables contribute to performance in NAPLAN tests. This statistical modelling is used to construct the SEA measure.

The SEA measure is based on two alternative data sources:

- Information relating to parent/guardian occupation, school education, non-school education and language background obtained from student enrolment records;
- Information relating to education and occupation variables corresponding to area level data from the Australian Bureau of Statistics (ABS) Census of Population and Housing.

Throughout this report, the parent/guardian background data obtained from enrolment records is referred to as 'direct parent data' and the Census data is referred to as 'indirect parent data'.

Indirect parent data are derived from the set of education and occupation variables used to develop the ABS Socio-Economic Indexes For Areas: Index of Education and Occupation (SEIFA: IEO). The indirect data are derived by relating student residential addresses to the smallest unit for Census data, which for the 2011 Census is Statistical Area Level 1 (SA1), and then extracting the information on nominated education and occupation variables from the corresponding SA1 units.

The final ICSEA is constructed by combining student-level SEA measures with school level indicators of educational advantage: geographical remoteness of school, as determined by the Accessibility/Remoteness Index of Australia (ARIA), and Aboriginal and Torres Strait Islander enrolments (ATSIE).

The model for calculation of ICSEA is based on the following formula:

$$\text{ICSEA} = \text{SEA} + \text{ARIA} + \text{Per cent ATSIE}$$

In 2010 and 2011, the ICSEA formula included a language background variable. Subsequent investigations showed that this variable is particularly susceptible to variations in the pattern of missing data and that it correlates highly with variables already in the model. Consequently, ACARA has discontinued the practice of adjusting ICSEA values for the disadvantaged language background other than English (LBOTE) difference coefficient in 2012 and will address language background as an SEA factor in its long-term research program.

Operationally, the generation of ICSEA includes three stages. Stage 1 involves the construction of an overall measure of school performance using the technique of factor analysis. In Stage 2 the technique of regression analysis is used to derive an equation describing the relationship between a range of community variables and the school performance measure. In Stage 3 this equation is used to construct the ICSEA scale.

The ICSEA scale has a mean of 1000 and a standard deviation of 100. ICSEA values range from around 500 (representing extremely educationally disadvantaged backgrounds) to about 1300 (representing schools with students with very educationally advantaged backgrounds).

To calculate ICSEA in 2012, ACARA has aligned the definition of the usable direct parent data with that in use to determine the distribution of students in SEA quartiles for *My School* reporting. Specifically, for a parent/guardian response to be deemed to be a usable record, there would need to be no missing responses to three parental educational and occupational variables.

Stage 1: Constructing the school performance scale

A school performance scale was constructed using 2009 NAPLAN data.

A primary performance scale was constructed using school mean scores for:

- Year 3 reading;
- Year 3 numeracy;
- Year 5 reading; and
- Year 5 numeracy.

A junior secondary performance scale was constructed using school mean scores for:

- Year 7 reading*;
- Year 7 numeracy*;
- Year 9 reading; and
- Year 9 numeracy.

The sets of primary and secondary NAPLAN means produced strong factors that explain 86.1% of the variance in the sets of means used to construct them.

A single performance scale was constructed from the separate primary and secondary scales by standardising the two scales and merging them. In this combined performance scale, each school's overall performance is expressed in terms of the number of standard deviations above or below the national mean: the primary mean for primary schools and the secondary mean for secondary schools.

* For jurisdictions that include Year 7 in primary school, performance scales for junior secondary schools were based on Year 9 results only. This has a negligible impact on the modelling, as the relative weights of the Year 7 and Year 9 means are very similar.

Stage 2: Constructing the SEA measures

2.1 Direct SEA scale

Direct parent data were available for students enrolled in pre-Year 1 to Year 12 in government schools and most non-government Catholic systemic schools, through the student background data collection (SBD). For some non-government systemic schools and most independent schools direct data were only available for students who participated in NAPLAN. For these schools, and for states/territories with Year 7 in secondary schools, data were available for students in Years 3, 4, 5 and 6 for primary schools and Years 7, 8, 9 and 10 for secondary schools. For Queensland, South Australia and Western Australia, data were available for Years 3 to 7 for primary schools and Years 8, 9 and 10 for secondary schools.

2.2 Student enrolment records

When enrolling a child in school, parents/guardians in all jurisdictions and sectors are asked which of the following categories best describes their occupation:

- Senior management in large business organisation, government administration and defence, and qualified professionals
- Other business managers, arts/media/sportspersons and associate professionals
- Tradespeople, clerks and skilled office, sales and service staff
- Machine operators, hospitality staff, assistants, labourers and related workers
- Not in paid work in last 12 months.

Parents/guardians are also asked which of the following four options best describes the school education level they achieved (What is the highest year of primary or secondary school completed?):

- Year 12 or equivalent
- Year 11 or equivalent
- Year 10 or equivalent
- Year 9 or equivalent or below.

Parents/guardians are also asked which of the following four options best describes their non-school education status (What is the level of the highest qualification completed?):

- Bachelor degree or above
- Advanced diploma/Diploma
- Certificate I to IV (including trade certificate)
- No non-school qualification.

Parents are also asked to indicate whether they speak a language other than English at home and if so, which ones.

2.3 Determining the composition of the direct parent data variable set

School-level occupation and education variables were constructed by dividing the number of parents/guardians in each response category by the number responding to the relevant question. For example, the school 'Professional' variable was constructed by dividing the number of respondents indicating that this was their occupation category by the number responding to one of the five occupation categories.

Even though the parent background data are collected at enrolment and are unlikely to be updated during the time that a student is enrolled in a school, the data should remain reasonably accurate. The school education level of parents will only change for the relatively few parents/guardians who undertake further secondary-level schooling through TAFE or an equivalent. The non-school education level will only change for the relatively small proportion of parents/guardians who undertake formal post-school education. Although many parents are likely to change jobs during the time that their children are enrolled in a school, they are likely to remain within the same occupation category.

The one variable which may change is the 'Unemployed' variable. Many parents re-enter the workforce during the time that their children are enrolled in a school. This is particularly so for women who have been full-time carers of preschool-aged children. Accordingly, the unemployed variable has not been used in the construction of the ICSEA. If some parents do move into the workforce this will also have a small effect on the other occupation variables; however, there is no way of predicting what this effect will be.

2.4 Simplifying direct parent data variable set

The data were used to construct 12 direct parent data school-level variables for inclusion in the analyses: four occupation variables, four school education variables and four non-school education variables.

Concerns have been expressed about the complexity of the equation to construct the SEA measure and about the degree of collinearity among 12 direct data variables. Accordingly, analyses were carried out to explore the feasibility of simplifying the equations for constructing both the direct and indirect parent data scales to be used in the construction of the ICSEA in 2010. Table 1 reports the results of these analyses for SEA calculated using direct data. Regression analyses were conducted using a 'stepwise' approach with $p_{in} = 0.05$ and $p_{out} = 0.10$. The correlations between the variables and the school performance measure are included for easy comparison.

The first solution reported is the maximum variance solution which includes all variables that make a statistically significant contribution to the explained variance. A widely accepted convention for detecting excessive collinearity amongst variables in a regression solution is that the variance inflation factor (VIF) for each of the variables should be less than 10.

The second solution reported in each table is the solution which provided the greatest explained variance but which has the regression weights (Betas) in the same direction as the correlation and has VIFs less than 10 for all variables. The correlation coefficient from such a solution was then used to calculate SEA based on direct data.

The correlation coefficients derived from this simplified set of variables were used to construct direct parent SEA in 2010 and 2011 and consequently were also used in 2012 ICSEA calculations. They are provided in the Appendix. Ellipses in the Solution 2 columns denote the reason for the variable's exclusion.

Table 1. Analyses for SEA calculated using direct data

Variables	Correlation	Maximum variance solution (EV=63.2%)		Solution 2 (EV=62.7%)	
		Beta	VIF	Beta	VIF
professional (O1)	0.646	-0.107	6.544	O	
associate professional (O2)	0.558	0.145	2.081	0.1	2.051
skilled non-professional (O3)	-0.133	-0.030	2.033	-	2.107
low skilled (O4)	-0.632	-0.037	4.228	N	
Year 12 or equivalent (SE4)	0.703	NS		O	
Year 11 or equivalent (SE3)	-0.244	0.103	1.292	O	
Year 10 or equivalent (SE2)	-0.529	NS		-	2.357
Year 9 or equivalent or below (SE1)	-0.521	NS		-	2.474
bachelor degree or above (NSE7)	0.714	0.772	8.937	0.3	9.018
advanced diploma/Diploma (NSE6)	0.539	0.167	1.976	0.0	2.781
certificate I to IV (NSE5)	-0.328	0.118	3.116	O	
no non-school qualification	-0.728	NS		-	9.545

2.5 Indirect SEA scale

Government and Catholic school students' residential addresses, used to construct the indirect parent data, are provided to ACARA by jurisdictions and Catholic education authorities. For independent school students, ACARA uses information provided by the ABS about which census collection unit individual students belong.

To calculate the 2012 SEA based on indirect data, ACARA has used an education and occupation variable as determined by the 2011 ABS Census of Population and Housing. The education and occupation variables in the 2011 ABS Census have generally remained the same as those in the 2006 ABS Census that were used to construct the indirect parent SEA in previous calculations of ICSEA.

The 2011 ABS Census SA1 data release unit available to ACARA is considerably smaller than that used in the past. Consequently the process of simplification of variables to determine the indirect parent data has been replicated using 2011 Census data aggregated at the SA1 level. The full list of variables and their abbreviations is provided in the Appendix.

This initial set of 32 SEIFA: IEO variables (excluding the percentage of Aboriginal and Torres Strait Islander people) was subject to the same regression analysis and the reduction process applied to simplify the direct parent data variable set, as described above.

Table 2 contains an overview of the model selection and the final proposed set of variables for the calculation of indirect SEA. Ellipses in the Solution 2 columns denote the reason for the variable's exclusion. The correlation coefficients for the simplified indirect parent data variables, including their full names, are provided in the Appendix.

Table 2. Model selection and proposed variables for the calculation of indirect ICSEA

2011 SEIFA Variables	Maximum variance solution (EV=54.6%)		Solution 2 VIF<10 (EV=46%)		
	Correlation	Beta	VIF	Beta	VIF
emp_p	0.29	0.39	27.68		O
FMCF_one_parent_p	-0.4	-0.16	1.86	-0.17	1.52
high_income_p	0.46	-0.13	4.3	O	
HSCP_year08_p	-0.4	-0.11	3.12	-0.2	2.13
HSCP_year09_p	-0.35	0.19	4.41	O	
HSCP_year12_p	0.55	0.62	12.82		O
notlf_p	-0.18	0.3	14.88		O
OCCP_clerical_admin_workers_p	0.25	-0.16	4.28	O	
OCCP_community_workers_p	-0.25	-0.16	1.75	-0.18	1.28
OCCP_labourers_p	-0.51	-0.1	2.82	-0.11	1.83
OCCP_manager_p	0.23	-0.1	6.31	O	
OCCP_sales_workers_p	0.14	0.07	2.04	0.13	1.54
OCCP_technician_p	-0.19	-0.16	5.55	-0.2	1.45
QALLP_4AD_Diploma_p	0.55	0.08	3.41	0.22	2.53
QALLP_51Certificate_III_IV_p	-0.32	0.13	6.66	O	
QALLP_52Certificate_II_I_p	-0.24	-0.05	1.13	-0.06	1.1
QALLP_71NoEducation_p	-0.14	-0.07	1.98	-0.04	1.34
unemp_p	-0.3	-0.04	2.39	-0.04	1.58

2.6 Aligning direct and indirect SEA scales

The following process was used to re-scale the indirect data scale to align it with the direct data scale:

1. The median of each decile of the indirect data scale was set at the same value as the corresponding median on the direct data scale.
2. The indirect scale values between medians were adjusted such that they retained their same relative position between the medians on the adjusted and unadjusted scales.
3. Values below the Decile 1 median were adjusted by using the Decile 1/2 adjustment factor; values above the Decile 10 median were adjusted using the Decile 9/10 adjustment factor.

Table 3 shows the scaling factors between each pair of medians. These were calculated by dividing the differences between the adjacent medians on the direct scale by the differences between the corresponding medians on the indirect scale.

Using these transformations, the indirect data SEA scale was equated to the direct data SEA scale. The resulting transformed indirect SEA measures were then used in further modelling.

Table 3. Scaling factors used to re-scale indirect scale values between adjacent decile medians

Scale section	Scaling factor
Below Decile 1 median	0.96
Decile 1 to Decile 2	0.96
Decile 2 to Decile 3	1.12
Decile 3 to Decile 4	1.35
Decile 4 to Decile 5	1.30
Decile 5 to Decile 6	1.46
Decile 6 to Decile 7	1.44
Decile 7 to Decile 8	1.36
Decile 8 to Decile 9	1.47
Decile 9 to Decile 10	0.93
Above Decile 10 median	0.93

Stage 3: Constructing the ICSEA

3.1 Inclusion of school variables

As in previous ICSEA calculation cycles, the quadratic and cubic variants of the SEA measure—the school percentage of Aboriginal and Torres Strait Islander enrolments and the school ARIA—were added to a regression model to produce the raw ICSEA values. This process was done separately for direct and indirect SEA measures.

3.2 Determining the source of the reported ICSEA

To further align the process of determining which data source will be used to determine the reported ICSEA, and upholding the original intention of ICSEA to be calculated using complete and reliable direct student background data, ACARA has applied the following rules to the 2012 ICSEA calculations:

1. Where sufficient direct parental data from enrolments exist, ICSEA is calculated by regressing NAPLAN performance on direct parent education and occupation variables, and on school-level ATSIE and ARIA variables.
2. Where sufficient direct parental data from enrolments do not exist, ICSEA is calculated by regressing NAPLAN performance on indirect data, where students' residential addresses inform proxy variables for parent education and occupation, drawing on such data available from the 2011 ABS Census.
3. The ICSEA value generated in step 1 or 2 above (as appropriate) is added to the ICSEA value published for 2011, and the result is divided by two.

3.3 Data sufficiency rules

For a school to qualify for direct parent data calculation it must have a sufficient number of available records. The number of required records is determined for each school depending on its size and the source of the direct parent data. Table 4 below shows rules and functions used to calculate such a number of required records. These rules were determined by the ICSEA Expert Advisory Group. If the number of available records is greater than the minimum required number of records the school is deemed to have the sufficient data for the calculation of ICSEA based on the direct parent data.

Table 4. Rules and functions to calculate minimum sufficient number of school records

Direct parent data source	Total enrolments (TE) for all available year levels	Formula for determining minimum required records (RR)
NAPLAN	less than or equal to 15	$RR = 0.9 * TE$
	greater than 15 but less than or equal to 110	$RR = 5 * \text{square root}(TE)$
	greater than 110	$RR = 0.33 * TE$
SBD	less than or equal to 30	$RR = 0.9 * TE$
	greater than 30 but less than or equal to 225	$RR = 5 * \text{square root}(TE)$
	greater than 225	$RR = 0.33 * TE$

3.4 Comparison of distributions of 2012 and 2011 ICSEA

Figure 1 shows the distribution of scores percentages for 2012 ICSEA plotted over the distribution of those reported in 2011.

The concentration of schools that have ICSEA values equal or close to the scale mean of 1000 appears to be somewhat higher in the 2012 distribution. However, as can be seen, the two distribution curves have very similar shapes and ranges, which indicates that there are no significant systemic movements in the distribution of 2012 ICSEA values.

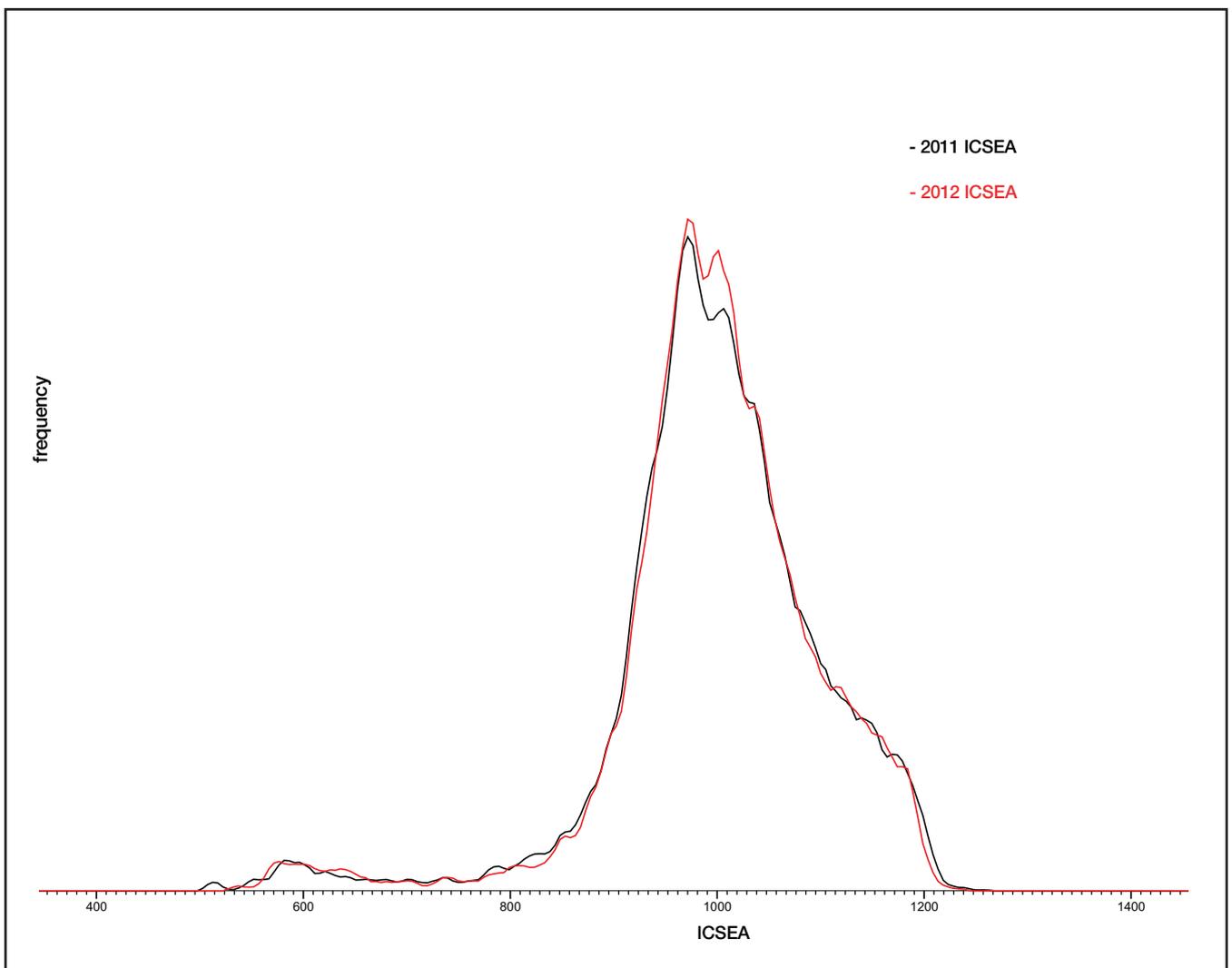


Figure 1. Distribution of 2011 and 2012 ICSEA values

Appendix

List of modelled 2011 SEIFA Variables

The variables below refer to the percentage of people in the data collection unit. For the school education variables and qualification variables this refers to people aged 25 years and over.

Table 5. Modelled 2011 SEIFA Variables

Short form	Description
OCCP_Manager_p	Managers
OCCP_Professionals_p	Professionals
OCCP_Technician_p	Technicians and Trades Workers
OCCP_community_workers_p	Community and Personal Service Workers
OCCP_clerical_admin_workers_p	Clerical and Administrative Workers
OCCP_sales_workers_p	Sales Workers
OCCP_mc_operators_drivers_p	Machinery Operators and Drivers
OCCP_labourers_p	Labourers
QALLP_1PG_Degree_p	Postgraduate Degree Level
QALLP_2GD_GC_level_p	Graduate Diploma and Graduate Certificate Level
QALLP_3BD_level_p	Bachelor Degree Level
QALLP_4AD_Diploma_p	Advanced Diploma and Diploma
QALLP_50Certificate_p	Certificate Level, nfd
QALLP_51Certificate_III_IV_p	Certificate III and IV Level
QALLP_52Certificate_II_I_p	Certificate I and II Level
QALLP_6School_Education_p	School Education Level
QALLP_71NoEducation_p	No educational attainment
HSCP_year12_p	Year 12 or equivalent
HSCP_year11_p	Year 11 or equivalent
HSCP_year10_p	Year 10 or equivalent
HSCP_year09_p	Year 9 or equivalent
HSCP_year08_p	Year 8 or equivalent or below
HSCP_no_school_p	No school education
FMCF_no_children_p	Family with no dependent offspring
FMCF_with_children_p	Family with dependent offspring
FMCF_one_parent_p	One-parent family with no dependent offspring
FMCF_other_parent_p	Other family
notlf_p	Not in the labour force
unemp_p	Unemployed adults
emp_p	Employed adults
high_income_p	High income
low_income_p	Low income

Direct parent data SEA equation

The regression equation used for calculation of direct SEA is provided below.

$$SEA\ direct = -0.18040638014807442$$

+0.0156748706921854 * Percentage of associate professionals

-0.00344552394075507 * Percentage of skilled non-professionals

-0.00615701087665729 * Percentage of parents with Year 10 or equivalent

-0.00526790993140476 * Percentage of parents with Year 9 or equivalent or below

+0.0190796607989309 * Percentage of parents with bachelor degree or above

+0.0145078895835201 * Percentage of parents with Advanced diploma/Diploma

+0.0113868380078018 * Percentage of parents with no non-school qualification

Indirect parent data SEA equation

The final regression equation for indirect data for 2012 ICSEA, with the full name of the variables in the simplified indirect data set, is provided below.

$$SEA\ indirect = 1.350803494021600$$

-3.125607989026660 * Percentage of families that are one parent families with dependent offspring only

-5.50907192610759 * Percentage of people 25 years and over who did not attend school beyond year 8 or below

-12.7745766939491 * Percentage of employed people with an occupation as a Community and Personal Services Worker

-4.49566005219386 * Percentage of employed with an occupation as a Labourer

+11.8074329016578 * Percentage of employed people with an occupation as a Sales Worker

-7.82672159268118 * Percentage of employed people with an occupation as a Technician or Trade Worker

+9.5445810085416 * Percentage of people aged 25 years and over with a diploma qualification

-27.7727270246363 * Percentage of people aged 25 years and over with a Certificate I or II qualification

-2.68133428194222 * Percentage of people 25 years and over with no non-school educational attainment

-3.24206211202633 * Percentage of unemployed adults

acara AUSTRALIAN CURRICULUM,
ASSESSMENT AND
REPORTING AUTHORITY

© Australian Curriculum, Assessment and Reporting Authority 2013