AUSTRALIAN CURRICULUM, ASSESSMENT AND REPORTING AUTHORITY

Level 10, 255 Pitt Street, Sydney NSW 2000
Phone 1300 895 563 **Fax** 02 8098 3103
Email info@acara.edu.au www.acara.edu.au

Report on the generation of the 2010 Index of Community Socio-Educational Advantage (ICSEA)

Dr Geoffrey Barnes

Parent socioeconomic data

The parent Socio-Educational Advantage (SEA) scale used in the construction of the 2010 Index of Community Socio-Educational Advantage is based on two alternative data sources:

- Information relating to parent occupation, school education, non-school education and language background obtained from student enrolment records
- Australian Bureau of Statistics (ABS) census data.

Throughout the report the parent background data obtained from enrolment records is referred to as '**direct parent data**' and the census data is referred to as '**indirect parent data**'.

Direct parent data was available for students enrolled in Kindergarten to Year 12 in government schools and most non-government systemic schools. For some non-government systemic schools and most Independent schools direct data was only available for students who participated in NAPLAN in 2009 and 2010. So, for these schools, and for states with Year 7 in secondary schools, data was 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 was available for Years 3 to 7 for primary schools and Years 9 and 10 for secondary schools.

Not all states and sectors provided updated address data for the generation of indirect parent data. Where 2010 address data was not available the most recent available data was used.

The construction of the ICSEA involves two stages. The first stage involves the construction of an overall measure of school performance using the technique of factor analysis. In stage two 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. This equation is then used to construct the ICSEA.

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
- Year 5 numeracy.

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

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

(* For jurisdictions which 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.)

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

A single performance scale was then constructed from the separate primary and secondary ones 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.

All factor analyses and regression analyses were carried out with schools with combined Year 3/Year 5 or Year 7/Year 9 cohorts of 20 or more students. The relationship between school average outcomes and community factors is often much weaker for small schools because they are much more susceptible to the influence of small numbers of students achieving at the top or bottom of the academic spectrum. The relationship between outcomes and community factors for small schools is not necessarily indicative of the general relationship between these variables.

Parent background data

When enrolling a child in school parents in all jurisdictions and sectors are asked which of the following five options 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
- Tradesmen/women, 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.

For convenience these five categories are referred to throughout the report as professional, semi-professional, skilled non-professional, low-skilled and unemployed.

Parents are also asked which of the following four options best describes the school education level they achieved.

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

Parents are also asked which of the following four options best describes their non-school education status.

- 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 one.

Direct parent data variables

Jurisdictions and sectors provided data for one or both parents depending on availability.

School-level occupation and education variables were constructed by dividing the number of parents 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 parents indicating that this was their occupation category by the number of parents providing a response to one of the five occupation categories.

Even though the parent background data is collected at enrolment and is unlikely to be updated during the time that a student is enrolled in a school it should remain reasonably accurate. The school education level of parents will only change for the very few parents that 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 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 pre school-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 but there is no way of predicting what this effect will be.

The data was 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.

Two alternative sets of variables were constructed:

- Combined parent variables
- Optimum parent variables.

The 'Combined parent' variables were constructed by adding the number of first and second parents in each response category and dividing by the total number of first and second parents responding to the relevant question.

The 'Optimum parent' variables were constructed by taking the higher skilled occupation category, the higher school education level and the higher non-school education level for each pair of parents, then calculating the school level percentages as above.

Population estimates and confidence Intervals

In most cases not all the school parent population provided responses to the questions about their occupation and education status and it was necessary to estimate the proportions in each category using the responses provided. The parents responding to the three questions were assumed to be random samples of the parent population and the sample proportions were taken as estimates of the population proportions.

It is also possible to calculate confidence intervals, for a given level of confidence, around the estimates of the population proportions. For example, if the proportion of a sample of parents in a particular category is 27% and the 95% confidence interval is calculated to be 3% we can say with 95% confidence that the proportion of all the parents in the school in the category is between 24% and 30%.

The standard error of an estimate of a population proportion based on sample data is given by the formula:

$$\sigma = \sqrt{\frac{\theta(1-\theta)(Np-N)}{N \qquad (Np-1)}}$$

Where

σ = the standard error of the proportion
θ = the proportion of the population in the category
$1-\theta$ = the proportion of the population not in the category
Np = the size of the population
N = the size of the sample drawn from the population

The population proportions in each category are unknown and it is necessary to assume that they are equal to the sample proportions.

The confidence interval for the estimate of a proportion of a population in a given category is equal to the standard error of the proportion multiplied by the z-score for the specified level of confidence.

$$CI = \sigma . z$$

A confidence level of 95%, a widely accepted convention, has been adopted for calculating confidence intervals. The relevant z-score for a 95% confidence interval is 1.96 and the formula for calculating confidence intervals becomes:

$$CI = 1.96 * \sqrt{\frac{\theta(1-\theta)(Np-N)}{N \qquad (Np-1)}}$$

The equation above relies on the assumption that the sampling distributions of the proportions are approximately normal. This assumption becomes less accurate as sample sizes decrease and the proportion of the population meeting the criterion differs from 50%. It is generally recommended that the formula for the standard error of a proportion should only be used when $N\theta$ or $N(1-\theta)$, whichever is the smaller, is less than 5. Thus if the proportion of interest is 0.2 (20% of the population) a sample size of at least 25 is required (0.2 * 25 = 5).

The response rates for the three questions varied considerably from state to state and from school to school. Table 3 shows the average school response rate across all schools and across schools in each state and sector, based on a response from at least one parent per family.

Where a response was provided to the school education question but no response was provided to the non-school education question, the parent was assumed to have had no non-school education.

Table 3: Average school percentage response rates

	Occupation	School education	Non-school education
All schools	77.9	80.9	74.0
ACT			
• Government	89.5	98.0	94.7
• Non-government	87.0	79.2	69.4
New South Wales			
• Government	75.9	83.8	72.9
• Non-government	87.2	85.6	81.4
Northern Territory			
• Government	36.5	38.8	33.7
• Non-government	42.0	42.4	40.4
Queensland			
• Government	78.7	81.4	73.8
• Non-government	75.1	72.2	70.3
South Australia			
• Government	58.1	76.8	67.9
• Non-government	83.2	81.6	74.6
Tasmania			
• Government	90.1	95.5	88.7
• Non-government	81.8	73.9	69.2
Victoria			
• Government	98.6	94.0	89.0
• Non-government	82.8	81.6	78.4
Western Australia			
• Government	53.7	59.1	50.5
• Non-government	82.3	84.3	81.8

The size of the confidence intervals around the population estimates are determined by the school response rate in conjunction with the size of the school population. The response rate required to obtain estimates with confidence intervals of a given size increases sharply as the size of the school population decreases. Table 4 shows how the required response rates for schools of different sizes vary for a 95% confidence level, a population proportion of 25% and a confidence level of 3%.

Table 4: Required response rates for a 95% confidence level, a population proportion of 25% and a confidence level of 3%.

School population	Required sample	Required response rate
25	24	96%
50	47	94%
75	69	92%
100	89	89%
150	126	84%
200	160	80%
500	308	62%
1000	445	45%

The data in Table 4 illustrates that for medium to large schools the required response rates are quite low and well within what is presently being achieved by most schools. However, it is often difficult for small schools to achieve the required response rates.

The direct parent data for many non-government schools related only to parents of students who were NAPLAN candidates in 2009 and 2010. In calculating the confidence intervals for these schools the students in the NAPLAN cohorts were assumed to be the entire school population. If the actual school populations were used the maximum possible response rate these schools could achieve, assuming yearly cohorts of equal size, would be 66%. This would make it virtually impossible for small and medium-sized schools to achieve acceptable confidence intervals. Note that this assumption does not affect the population estimates themselves, just the confidence intervals around them and, as is shown below, whether the direct or indirect parent data is used to calculate the school ICSEA values.

The accuracy of the school population estimates is critical in assessing the relative merits of the direct and indirect parent data for calculating the ICSEA. Accordingly, for each school the confidence intervals for each of the 12 population estimates were calculated and the average confidence interval was determined. For a given school the response rates to the three questions may vary, as will the proportions themselves, so the confidence intervals will vary. However, the average confidence interval provides a convenient indication of the overall accuracy of the school's data.

Table 5 shows the cut-off values for the average confidence interval deciles for the two alternative sets of parent variables. Table 5 shows, for example, that 10% of schools had an average confidence interval less than 0.8% for the combined parent data; for 10% of schools we can assume with 95% confidence that, on average, the estimates of the population proportions provided by the combined parent data are accurate to within 0.8%. Similarly we can assume that the estimates for 20% of schools are within 1.0%.

Table 5: Confidence interval cut-off values

Decile	Confidence interval cut-off value	
	Combined	Optimum
1	0.0 to 0.8	0.0 to 0.5
2	0.8 to 1.0	0.5 to 0.9
3	1.0 to 1.3	0.9 to 1.2
4	1.3 to 1.5	1.2 to 1.6
5	1.5 to 1.8	1.6 to 1.9
6	1.8 to 2.2	1.9 to 2.4
7	2.2 to 2.7	2.4 to 3.1
8	2.7 to 3.6	3.1 to 4.2
9	3.6 to 5.7	4.2 to 6.6
10	5.7 to 34.0	6.6 to 39.7

Comparing direct parent data variable sets

Table 6 reports the school average proportions for the different occupation and education categories for the four direct parent data variable sets. The averages relate to data aggregated to the school level not to aggregated national data. Table 7 reports the correlations between the variables and school performance.

Table 6: School average proportions of parents in occupation and education categories

	Combined	Optimum
Occupation variables		
• Professional (O1)	14.9	21.5
• Associate professional (O2)	20.8	25.3
• Skilled non-professional (O3)	21.9	22.6
• Low skilled (O4)	21.3	19.2
School education variables		
• Year 12 or equivalent (SE4)	51.3	63.0
• Year 11 or equivalent (SE3)	13.9	13.4
• Year 10 or equivalent (SE2)	26.5	18.7
• Year 9 or equivalent or below (SE1)	8.2	4.9
Non-school education variables		
• Bachelor degree or above (NSE7)	20.0	29.5
• Advanced diploma/Diploma (NSE6)	11.4	15.2
• Certificate I to IV (NSE5)	28.2	33.4
• No non-school qualification (NSE8)	40.5	21.9

Table 7: Correlations between proportions of parents in occupation and education categories and school performance

	Combined	Optimum
Occupation variables		
• Professional (O1)	.646	.659
• Associate professional (O2)	.558	.478
• Skilled non-professional (O3)	-.133	-.330
• Low skilled (O4)	-.632	-.609
School education variables		
• Year 12 or equivalent (SE4)	.703	.719
• Year 11 or equivalent (SE3)	-.244	-.387
• Year 10 or equivalent (SE2)	-.529	-.570
• Year 9 or equivalent or below (SE1)	-.521	-.464
Non-school education variables		
• Bachelor degree or above (NSE7)	.714	.721
• Advanced diploma/Diploma (NSE6)	.539	.214
• Certificate I to IV (NSE5)	-.328	-.548
• No non-school qualification (NSE8)	-.728	-.632

Each set of parent variables was regressed on the school performance scale. Because the explained variance is influenced by the accuracy of the school population estimates, analyses were conducted with groups of schools with increasing average confidence interval cut-offs. Table 8 reports the variance in the school performance measure explained by the different sets of variables and groups of schools.

Table 8: Variance explained by sets of parent variables with different average confidence intervals

Variable Set	**Variance explained**						
	CI<=0.5	CI<=1.0	CI<=1.5	CI<=2.0	CI<=2.5	CI<=3.0	CI>3.0
Combined parent variables*	72.6% (214)	65.1% (1627)	65.3% (3435)	64.1% (4763)	63.9% (5643)	63.2% (6144)	52.1% (883)
Optimum parent variables	70.2% (583)	65.6% (1848)	64.0% (3120)	63.6% (4211)	63.3% (5039)	62.8% (5598)	52.4% (1429)
Indirect parent data	54.6% (6960)						

* It was necessary to assume that the population was twice the number of students enrolled because the coding system does not distinguish between ‘not known’ rather than ‘missing’. It was not possible to adjust the population to account for single parent families.

The decision as to which of the parent data sets is the most suitable for the construction of the ICSEA has been based on the criterion that it should have the greatest explanatory power for the greatest number of schools. Table 8 shows that the combined parent variables generally explain a greater proportion of the variance in the performance measure than the optimum parent variables.

The bottom row of Table 8 reports the proportion of variance in the school outcome scale explained by the indirect parent data. The indirect data scale was constructed by regressing the census data variables on the school performance measure. One of these variables, the 'Percentage of people who identified themselves as being of Aboriginal or Torres Strait Islander origin' was omitted leaving 13 of the 14 original variables. This was done to construct an indirect parent data scale which was analogous to the direct data scales. The net impact of omitting this variable was negligible because, as will be reported below, its omission results in a substantial increase in the variance explained by the 'school ATSI enrolment' variable added in the second stage of the ICSEA construction.

Simplifying the direct and indirect parent data equations

Concerns have been expressed about the complexity of the equation used to construct the ICSEA in 2009 and about the degree of collinearity amongst the ICSEA variables. In 2009 several of the variables had signs in the opposite direction to their correlations. 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. Tables 9 and 10 report the results of these analyses. Regression analyses were conducted using a 'stepwise' approach with p in =.05 and p out = .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 variable weights from these solutions have been used to construct the direct and indirect parent data scales.

Table 9: Alternative regression solutions for direct parent data variables

	Correlation	Maximum variance solution (EV=63.2%)		Solution 2 (EV=62.7%)	
		Beta	**VIF**	**Beta**	**VIF**
Occupation variables					
• Professional (O1)	.646	-.107	6.544	O	
• Associate professional (O2)	.558	.145	2.081	.154	2.051
• Skilled non-professional (O3)	-.133	-.030	2.033	-.031	2.107
• Low skilled (O4)	-.632	-.037	4.228	NS	
School education variables					
• Year 12 or equivalent (SE4)	.703	NS		O	
• Year 11 or equivalent (SE3)	-.244	.103	1.292	O	
• Year 10 or equivalent (SE2)	-.529	NS		-.092	2.357
• Year 9 or equivalent or below (SE1)	-.521	NS		-.042	2.474
Non-school education variables					
• Bachelor degree or above (NSE7)	.714	.772	8.937	.364	9.018
• Advanced diploma/Diploma (NSE6)	.539	.167	1.976	.078	2.781
• Certificate I to IV (NSE5)	-.328	.118	3.116	O	
• No non-school qualification (NSE8)	-.728	NS		-.196	9.545

Table 10: Alternative regression solutions for indirect parent data variables

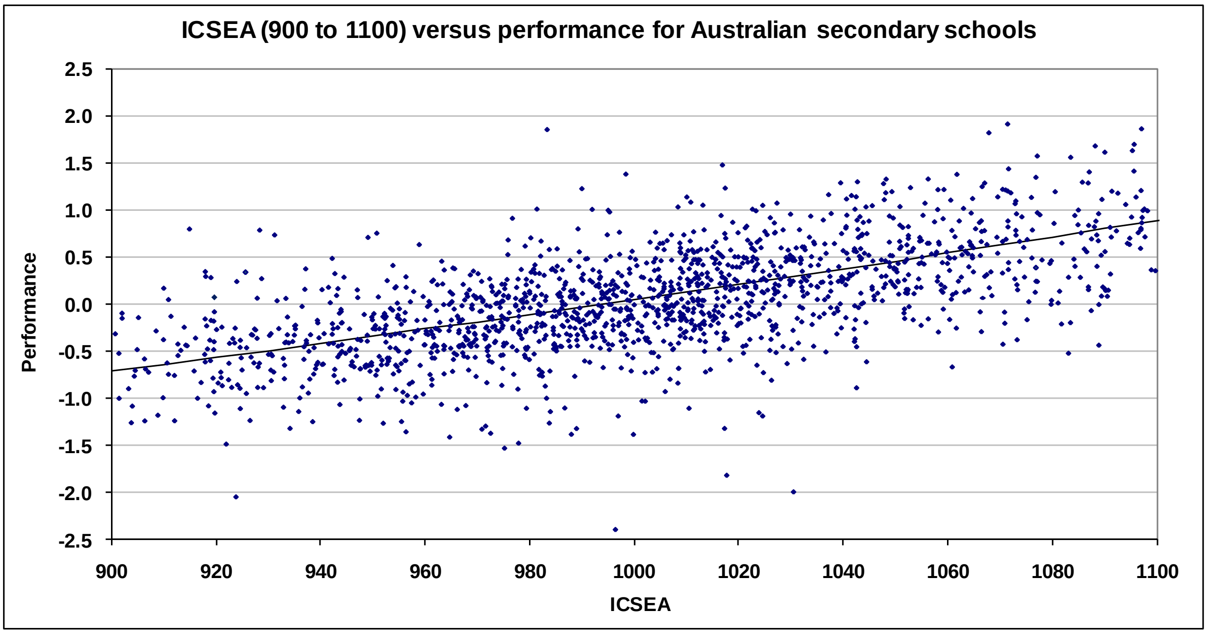
	Correlation	Maximum variance solution (EV=54.6%)		Solution 2 VIF<10 (EV=51.5%)	
		Beta	VIF	Beta	VIF
Income variables					
Percentage of people with annual household income between $13,000 and $20,799 (INC_LOW)	-.405	.070	9.656	O	
Percentage of people with annual household income greater than $52,000 (INC_HIGH)	.504	-.341	9.996	O	
Education variables					
Percentage of people aged 15 years and over with a certificate qualification (CERT)	-.275	-.320	10.280	-.123	1.433
Percentage of people 15 years and over with an advanced diploma or diploma qualification (DIP)	.583	NS		O	
Percentage of people 15 years and over with no post-school qualifications (NOQUAL)	-.629	-.448	14.808	-.142	6.366
Percentage of people 15 years and over whose highest level of schooling completed is Year 11 or lower (NOYEAR12)	-.570	.229	21.324	O	
Percentage of people 15 years and over who did not go to school (NOSCHOOL)	-.098	.038	2.102	O	
Employment variables					
Percentage of people (in the labour force) who are unemployed (UNEMP)	-.345	.068	2.561	O	
Occupation variables					
Percentage of employed people who work in a skill level 1 occupation (OCC_1)	.630	.163	9.381	O	
Percentage of employed people who work in a skill level 4 occupation (OCC_4)	-.341	NS		-.177	2.218
Percentage of employed people who work in a skill level 5 occupation (OCC_5)	-.555	-.105	2.619	-.091	2.277
Others					
Percentage of families that are one parent families with dependent offspring only (ONEPAR)	-.551	-.289	2.211	-.244	1.486
Percentage of occupied private dwellings with no internet connection (NONET)	-.528	-.407	8.235	-.278	5.127

Using direct or indirect parent data

For most schools two alternative sources of parent data are available; direct data (parent enrolment data) and indirect data (ABS census data). Criteria need to be established for determining, school by school, which of these data sources provides the more accurate measure of socio-educational advantage. It is argued above that the average of the confidence intervals around the estimates of the percentages of parents in the occupation and education categories provides a convenient indicator of the accuracy of this data and the results reported in Table 8 demonstrate that if the average confidence intervals are small enough the direct data provides a more accurate measure of socio-educational advantage than the indirect data. The issue becomes one of determining the point at which the direct data ceases to provide a more accurate assessment of socio-educational advantage than the indirect data.

The most appropriate way to explore this issue is to analyse the residuals (distances) of the data points about the regression lines produced by the direct and indirect data. Figure 1, taken from the 2009 ICSEA modelling report, shows the indirect data regression line between ICSEA values of 900 and 1100.

Figure 1: ICSEA versus performance for Australian secondary schools



Inherent in the logic underlying the construction of the ICSEA is the idea that the residuals are the result of variation in school effectiveness – the regression line describes the component of school performance which can be attributed to community factors and the residual represents the component which can be attributed to school practices. Realistically, however, a portion of each residual is the result of measurement error, either in measuring academic performance or socio-educational advantage. It follows, therefore, that reduced residuals are an indication of reduced measurement error.

By comparing the absolute magnitude of the residuals about the direct data regression line for groups of schools with different average confidence intervals with the absolute magnitude of the residuals about the indirect data regression line for all schools, it is possible to estimate the average confidence interval value at which the direct data becomes less accurate than the indirect data in describing community socio-educational advantage. Table 11 present the results of this analysis.

Table 11: Comparison of absolute residuals about direct data and indirect data regression lines

	Average of absolute residuals*					
	CI<=1.0	1.0<CI<=1.5	1.5<CI<=2.0	2.0<CI<=2.5	2.5<CI<=3.0	3.0<CI<=3.5
Direct data	.3779	.3750	.4041	.4477	.4809	.5085
Indirect data	.4812					

* The average of the absolute values is used because half the residuals are positive and half are negative. The average of the actual residuals is zero.

The average of the residuals for all schools about the indirect parent data regression line is .4812. As expected there is a general increase in the average of the residuals about the direct data regression line as the average confidence interval increases. The point at which the direct data becomes less accurate than the indirect data, the point at which the average residual for the direct data becomes greater than the average residual for the indirect data, occurs when the average confidence interval reaches about 3.0%. This has been used as the criterion for determining which of each school's alternative sets of parent data is used. Direct parent data has been used for schools where the average of the confidence intervals around the population estimates is equal to or less than 3.0% and indirect data has been used for schools with confidence intervals greater than this.

Table 12 shows the numbers and percentages of schools by state and sector with confidence intervals equal to or less than 3.0%; the numbers and percentages of schools that have their ICSEA values based on direct parent data.

Table 12: Numbers and percentages of schools with confidence intervals equal to or less than 3.0%

State and Sector	No of schools	Schools with average confidence interval less than 3.0%	
		Number	Percentage
ACT Government			
• Government	83	74	89.2%
• Non-government	43	38	88.4%
NSW Government			
• Government	2136	1504	70.4%
• Non-government	905	731	80.8%
Northern Territory			
• Government	149	46	30.9%
• Non-government	30	11	36.7%
Queensland			
• Government	1242	918	73.9%
• Non-government	470	310	66.0%
South Australia			
• Government	544	332	61.0%
• Non-government	198	149	75.3%
Tasmania			
• Government	187	180	96.3%
• Non-government	65	38	58.5%
Victoria			
• Government	1493	1338	89.6%
• Non-government	764	589	77.1%
Western Australia			
• Government	746	388	52.0%
• Non-government	294	220	74.8%
Total	9865	6976	70.7%

Note: Numbers do not add up to totals because some schools could not be allocated to a particular sector.

It was demonstrated earlier that a school's average confidence interval is determined by the size of its enrolment as well as its response rate; small schools need greater response rates to achieve similar confidence intervals to large schools - see Table 4. As expected, therefore, the majority of the schools with confidence intervals greater than 3.0% are small schools. Half of these schools have enrolments of less than 60 students.

Aligning direct and indirect parent data scales

There were 6789 schools that had:

- an average confidence interval for their direct parent data less than or equal to 3.0, and
- a value on the indirect parent data scale.

These 6789 'moderation schools' were divided into 10 groups of equal size (deciles) on the direct and indirect parent data scales and the medians of the sets of deciles were determined. (Note that a particular school is not necessarily in the same decile on the two different scales.

Table 13 reports the medians for the two sets of deciles and Figure 2 shows the relationship between them. Note that the scales are still in a roughly standardised form.

Table 13: Medians for indirect and direct parent data scale deciles

Decile	Indirect data scale	Direct data scale
1	-0.9642	-0.9820
2	-0.6148	-0.6973
3	-0.4545	-0.5101
4	-0.3133	-0.3380
5	-0.1705	-0.1527
6	-0.0238	0.0344
7	0.1578	0.2481
8	0.4165	0.5134
9	0.7591	0.8652
10	1.2647	1.3773

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

- 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.
- 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 (See Example 1 below).
- 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 Decile9/10 adjustment factor (See Example 2 below).

Table 14 below shows the adjustment 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. For example the adjustment factor between the medians for deciles 7 and 8 is:

Adjustment factor = (0.5124 – 0.2481) / (0.4165 – 0.1578) = 1.0255

Table 14: Adjustment factors used to re-scale indirect scale values between adjacent decile medians

Scale section	Scaling factor
Below Decile 1 median	0.8145
Decile 1 to decile 2	0.8145
Decile 2 to decile 3	1.1685
Decile 3 to decile 4	1.2189
Decile 4 to decile 5	1.2969
Decile 5 to decile 6	1.2762
Decile 6 to decile 7	1.1760
Decile 7 to decile 8	1.0255
Decile 8 to decile 9	1.0270
Decile 9 to decile 10	1.0128
Above decile 10 median	1.0128

Example 1: The re-scaled value for a school with a value of 0.3 on the unadjusted scale (between the Decile 7 and Decile 8 medians) would be

Value = 0.2481 + 1.0255*(0.3000 – 0.1578) = .3939

Example 2: The re-scaled value for a school with a value of 1.5 on the unadjusted scale (Above the Decile 10 median) would be

Value = 1.3773 + 1.0128*(1.5000 – 1.2647) = 1.6156

A combined parent data scale was then constructed. Schools with average confidence intervals less 3.0% were allocated their value from the direct parent data scale and the remainder were allocated their value from the re-scaled indirect parent data scale.

Table 15 reports the variance explained by the indirect parent data scale for all schools, the direct parent data scale for schools with average confidence intervals less than 3.0% and the combined parent data scale for all schools. The results demonstrate conclusively that using the direct parent data where possible substantially increases the explanatory power of the parent socioeconomic data.

Table 15: Variance explained by direct, indirect and combined parent data scales

	Explained variance
Indirect parent data scale	51.5%
Direct parent data scale	62.7%
Combined parent data scale	58.8%

Inclusion of school variables – construction of the ICSEA

As in 2009 the quadratic and cubic variants of the parent data scale, the ‘school percentage of Aboriginal and Torres Strait (ATSI) enrolments’ and the school Accessibility/Remoteness Index of Australia (ARIA) values were then added to produce the ICSEA scale. Table 16 shows the results of progressively including these additional variables.

Table 16: Additional and total variance explained as additional variables are added

Variables	Explained variance	
	Additional	Total
Parent data		58.8%
+ Percentage of ATSI enrolments	8.9%	67.7%
+ parent data squared	0.5%	68.2%
+ parent data cubed	<0.1%	68.2%
+ ARIA	<0.1%	68.2%

The additional variance explained by the ATSI enrolment variable (8.9%) is approximately twice as large as in the 2009 ICSEA scale. This is because the ABS variable ‘Percentage of people who identified themselves as being of Aboriginal or Torres Strait Islander origin’ was omitted from the calculation of the indirect data scale and there was no equivalent variable used in the construction of the direct data scale.

The inclusion of the quadratic component of the parent data scale increased the explained variance by 0.5% indicating that the relationship between parent socio-educational status and school performance is non-linear. The contribution of the remaining two variables was less than 0.1%.

The preliminary set of ICSEA values was scaled to a mean of 1000 and a standard deviation of 100.

Inclusion of a disadvantaged LBOTE adjustment

LBOTE students usually perform marginally better on average than their non-LBOTE colleagues. However, there is considerable variation in performance across the different language groups within the LBOTE community with some language groups being particularly disadvantaged. In response to community concerns, the 2010 ICSEA includes an adjustment for schools with students from these disadvantaged language groups.

Supplementary analyses show that parents of students in these disadvantaged language groups are likely to have lower school education levels than other LBOTE parents. Accordingly, an additional variable, the percentage of parents in the school community who were both LBOTE and who reported having a maximum school education level of Year 9 or equivalent was included in the calculation of the ICSEA. This additional variable is referred to as the ‘Disadvantaged LBOTE variable’.

The LBOTE adjustment was limited to schools with a confidence interval less than or equal to 3.0% around the Disadvantaged LBOTE variable – the target schools. Because the adjustment was limited to a subset of schools it could not be calculated by simply adding the LBOTE school education variable to the preliminary ICSEA scale which includes all schools, as was done with the ‘school percentage of ATSI enrolments' variable. The adjustment was calculated by carrying out separate regression analyses with the target schools with and without the Disadvantage LBOTE variable and calculating the difference. These differences were then subtracted from the preliminary ICSEA values of the target schools. The inclusion of this variable increased the explanatory power of the ICSEA by 0.1% for the target schools.

The Disadvantaged LBOTE adjustment factor is approximately -0.5 ICSEA points for each one per cent of disadvantaged LBOTE students in the school. The maximum possible reduction in the school ICSEA score is therefore approximately 50 points.

Comparison of ICSEA 2009 and 2010

Tables 17 and 18 provide a comparison of the variables used in the construction of ICSEA in 2010 and 2009.

Table 17: Variables used in the construction of ICSEA in 2010

Component	Data source	
Socio-educational information	For 71% of schools this comprises 7 variables constructed from data supplied directly by parents – see Table 9	For 29% of schools this comprises 6 variables constructed from estimates based on ABS census data – see Table 10
Proportion of ATSI enrolments	The proportion of Aboriginal and Torres Strait Islander students enrolled in the school as indicated in school enrolment records	
Accessibility/Remoteness	The school's value on the Accessibility/Remoteness Index of Australia (ARIA)	
Proportion of disadvantaged LBOTE students	The proportion of students from LBOTE families with parents having low school education levels as indicated in school enrolment records	

Table 18: Variables used in the construction of ICSEA in 2009

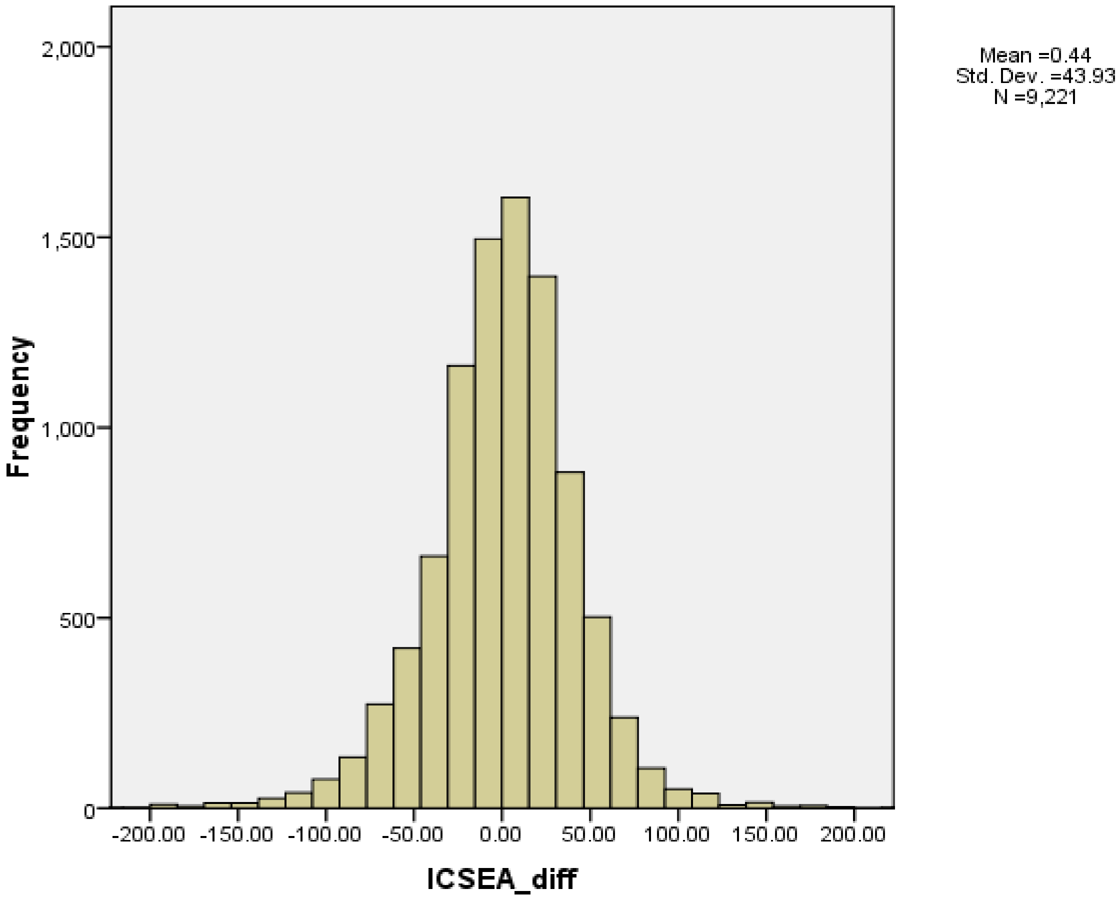
Component	Data source
Socio-educational information	Thirteen variables constructed from estimates based on ABS census data - the variables listed in Table 10 excluding the proportion of unemployed parents and including the proportion of Aboriginal families in the community.
Proportion of ATSI enrolments	The proportion of Aboriginal and Torres Strait Islander students enrolled in the school as indicated in school enrolment records
Accessibility/Remoteness	The school's value on the Accessibility/Remoteness Index of Australia (ARIA)

Impact of changes in the calculation of ICSEA

Table 19: ICSEA 2009/2010 difference deciles

Decile	Difference
1	-349.0 to -47.9
2	-47.9 to -32.0
3	-32.0 to -20.9
4	-20.9 to -11.6
5	-11.6 to -2.5
6	-2.5 to 6.8
7	6.8 to 6.9
8	16.9 to 29.5
9	29.5 to 49.5
10	49.5 to 428.3

Figure 2: Histogram of the differences between direct and indirect parent data ICSEA values



The data displayed above can be summarised as follows:

- The changes for the 20% of schools (deciles 5 and 6) will be within approximately 10 points on the ICSEA scale
- The changes for 20% of schools (deciles 4 and 7) will be between approximately 10 and 20 points.
- The changes for 20% of schools (deciles 3 and 8) will be between 20 and 30 points
- The changes for 20% of schools (deciles 2 and 9) will be between approximately 30 and 50 points.
- The changes for 20% of schools will greater than 50 points.

Comparing changes for different sectors

Concerns have been expressed in some quarters that the assumption of census collection district homogeneity results in an underestimation of ICSEA values for non-government schools. Table 18 shows the average differences for schools in the government, non-government systemic and non-government non-systemic sectors.

Table 20: Average difference between direct and indirect parent data ICSEA values by school sector

Sector	Average difference
Government	-10.2
Systemic schools	15.7
Non-systemic schools	24.3

2011 business rules applied to the 2010 ICSEA methodology

Rationale: Validity and stability of ICSEA values

Ideally, the Index of Community Socio-Educational Advantage (ICSEA) would be calculated using complete and reliable direct student background data for all students in all schools. However, analysis of the data provided for these ends indicates some shortcomings. While work is ongoing to address these shortcomings, the approved process used to develop ICSEA values in 2010 was used again in 2011 in order to maximise validity and stability.

Process: Determining the best possible fit between ICSEA and NAPLAN performance using the best available data

ACARA calculated each school's 2011 ICSEA using the approved 2010 methodology.

Where the difference between 2010 and 2011 values was not substantial, the weighted average of the 2010 and 2011 values is treated as the reported ICSEA value for 2011. The 2010 and 2011 ICSEA values were weighted according to the number of students for whom at least one parent had responded.

Where the difference between the two values was substantial, the ICSEA value reported is the value that is closer to the 2011 performance measure—either the 2010 ICSEA or the weighted average of the two ICSEA values—based on the number of student records used to generate both ICSEA values.

The difference in data quality between parental occupation and education ('direct') information and Australian Bureau of Statistics Census Collection District level ('indirect') data sources, however, suggest that a weighting approach would not be appropriate where an ICSEA value was generated from 'direct' data in 2010 but had reverted to 'indirect' data in 2011. Analysis conducted of the 2010 ICSEA methodology indicated an increase in the predictive power of ICSEA against NAPLAN performance from 59% based on the 'indirect' method to 68% when calculated through 'direct' data. As a result, 'direct' ICSEA values are used in 2011 rather than 'indirect' ICSEA values. The same rationale was applied where 'direct' ICSEA values were generated in 2011 for schools that had 'indirect' values in 2010.

Where the 2010 ICSEA value was developed through a combination of methodologies, ACARA adopted that value as the 2011 ICSEA value, unless otherwise specified by jurisdictional authorities.

Sensitivity: Defining 'substantial change' as 0.2 standard deviation

'Substantial change' is defined as change greater than 20 ICSEA points; that is, a change representing 0.2 of a standard deviation between the two values (given a standardised mean of 1000 and standard deviation of 100). This was modelled on data as they were received from schools, systems and sectors, and was established once all 2011 student background data were received.

The decision rules are summarised in the table below.

Category	Decision rule
a) 2010 & 2011 direct (and **no** substantial difference)	Use weighted average of 2010 and 2011 ICSEA
b) 2010 & 2011 indirect (and **no** substantial difference)	Use weighted average of 2010 and 2011 ICSEA
c) 2010 indirect but 2011 direct (regardless of difference)	Use 2011 ICSEA
d) 2010 direct but 2011 indirect (regardless of difference)	Use 2010 ICSEA
e) 2010 & 2011 direct (but there **is** substantial difference)	Use whichever of i) 2010 or ii) weighted average of 2010 and 2011 that best predicts NAPLAN performance
f) 2010 & 2011 indirect (but there **is** substantial difference)	Use whichever of i) 2010 or ii) weighted average of 2010 and 2011 that best predicts NAPLAN performance
g) Combined data sources used in 2010	Default to use 2010 ICSEA

Table 1: Decision rules summary for 2011 ICSEA