Technical paper

Index of Community Socio-Educational Advantage (ICSEA)

This brief paper provides a technical overview of the construction and use of the Index of Community Socio-educational Advantage (ICSEA) on the *My School* website to identify socio-educationally similar schools across Australia for the purposes of comparing the performance of schools.

Introduction

The Index of Community Socio-Educational Advantage (ICSEA) was developed specifically for the *My School* website as a means of identifying socio-educationally similar schools across Australia. It is tailored specifically towards educational outcomes, unlike more general measures of socio-economic status.

The first section of the paper summarises the method used to develop the index, the data used to develop it, and the results of various analyses to assess its validity.

The second section of the paper outlines the steps used to calculate a school's ICSEA value for the *My School* website and identifies limitations of the index. It also explains how the index is used on the website to compare the performance of schools.

1. Development of the index

The development of ICSEA was overseen by the Schools Reporting Working Group comprising senior officials from ACARA, State, Territory and Commonwealth education departments, and representatives from the non-government schools sector. The construction and modelling process was subsequently reviewed by an expert panel comprising the following persons:

Chair:
Professor Barry McGaw (Chair, ACARA)

Members:
Dr Geoff Barnes (NSW Department of Education and Training)
Mr John Firth (Chief Executive, VCAA)
Assoc Prof Shelley Gillis (Victoria University)
Mr Michael Long (Senior Research Fellow, Monash University)
Ms Julie McMillan (Fellow, ANU)

The research underpinning ICSEA was conducted by Dr Geoff Barnes.

Method

The best way to compare the academic performance of schools is to find groups of schools with students of similar abilities on commencing school. Unfortunately, no such measures of starting abilities are currently available nationally, so instead, attention focused on finding proxy measures that are highly correlated with student performance.

Research was undertaken to find a set of variables that best predicted student performance on the National Assessment Program – Literacy and Numeracy (NAPLAN) tests of Reading and Numeracy, and to use these to create an index that could be used to group schools that are 'statistically similar'.

The method used involved two stages. In stage 1, overall measures of school performance as indicated by NAPLAN results were constructed using factor analysis. These measures were used as the dependent variables for the second stage.

In stage 2, regression analysis was used to derive a mathematical equation describing the relationship between a range of community variables and school performance. The community variables were all derived by obtaining the addresses of students and linking them to Australian Bureau of Statistics data on Census Collection Districts (CCDs). The values for each of the parameters of the mathematical equation were then used to construct the Index of Community Socio-Educational Advantage (ICSEA).

Data

NAPLAN results, home addresses and Australian Bureau of Statistics (ABS) data were obtained for 9159 government and non-government schools across Australia, representing all schools except a very small number for which it was not possible to geo-code address data.

The ABS does not provide data for 3.2% of Census Collection Districts because of the unreliability of the data. These CCDs are mainly in remote and/or very disadvantaged areas. The data for these CCDs were not used in the development of ICSEA.

Results

Constructing overall measures of school performance

As part of stage 1, an overall performance scale was constructed for primary schools and selected year levels of secondary schools using mean scores from NAPLAN tests on Reading and Numeracy for Years 3, 5, 7 and 9.

The analysis produced a very strong single factor that explained 86.1% of the variance in the eight separate sets of mean scores.

Constructing the Index of Socio-Educational Advantage

There is universal agreement in the educational research literature that socio-economic factors play a major role in explaining educational outcomes. The Australian Bureau of Statistics (ABS) produces four indices of socio-economic status, the Socio-economic Indicators for Areas (SEIFA), which are intended for different purposes. However none of these has been designed specifically for use in educational contexts.

The SEIFA indices are calculated from a range of variables constructed from ABS census data. The second stage in the development of ICSEA was to identify that set of variables that best explained variation in the overall measures of school performance calculated in stage 1.

There are 35 variables in all that are used to construct the four SEIFA indices. Fifteen of these were selected on the basis of previous research as likely to be highly correlated with the overall performance measures.

The 15 variables and their correlation with the outcome measures calculated in stage 1 are summarised in Table 1. Two sets of Census Collection Districts (CCD) values for these 15 variables were used: one set for all households and one set for households with school-age children. The correlations for only one of the variables, 'percentage of people who do not speak English well', were below statistical significance. This variable was omitted from further analyses and was not included in the calculation of ICSEA.

The two sets of the remaining 14 SEIFA component variables, one set for 'all households' and one set for 'households with school-age children', were regressed on the school performance measures and the variable weights obtained from these analyses were used to construct scales of socio-educational advantage. The amount of variation in the performance measures explained by these scales was then determined and compared to that explained by the four SEIFA indices.

The proportion of variance explained by the 14 variables was much higher than that explained by the four ABS indices of socio-economic status, providing strong support for the construction of a measure specifically for use in educational settings. The results are displayed in Table 2.

The proportion of variance explained by the data for 'households with school-age children' was between one and two percentage points higher than that explained by the data for 'all households'. However, using only the information collected from households with school-age children greatly reduces the amount of data available for calculating school ICSEA values, so the decision was taken to make use of 'all households' data in computing ICSEA values.

Table 1: Correlations between selected SEIFA variables and overall school performance

Variable	Primary		Secondary	
	All households	Households with school-age children	All households	Households with school-age children
Income variables				
Percentage of people with annual household income between $13,000 and $20,799 (INC_LOW)	-.406**	-.519**	-.398**	-.537**
Percentage of people with annual household income greater than $52,000 (INC_HIGH)	.491**	.568**	.490**	.575**
Education variables				
Percentage of people aged 15 years and over with a certificate qualification (CERT)	-.246**	-.206**	-.198**	-.179**
Percentage of people 15 years and over with an advanced diploma or diploma qualification (DIP)	.588**	.516**	.552**	.527**
Percentage of people 15 years and over with no post-school qualifications (NOQUAL)	-.627**	-.693**	-.576**	-.663**
Percentage of people 15 years and over whose highest level of schooling completed is Year 11 or lower (NOYEAR12)	-.544**	-.587**	-.514**	-.575**
Percentage of people 15 years and over who did not go to school (NOSCHOOL)	-.109**	-.147**	-.172**	-.167**
Employment variables				
Percentage of people (in the labour force) who are unemployed (UNEMP)	-.342**	-.340**	-.420**	-.394**
Occupation variables				
Percentage of employed people who work in a skill level 1 occupation (OCC_1)	.623**	.678**	.618**	.677**
Percentage of employed people who work in a skill level 4 occupation (OCC_4)	-.338**	-.367**	-.284**	-.326**
Percentage of employed people who work in a skill level 5 occupation (OCC_5)	-.571**	-.517**	-.499**	-.471**
Others				
Percentage of people who do not speak English well (ENGPOOR)+	.036	-.003	-.020	-.037
Percentage of families that are one-parent families with dependent offspring only (ONEPAR)	-.549**	-.474**	-.563**	-.486**
Percentage of occupied private dwellings with no internet connection (NONET)	-.540**	-.702**	-.537**	-.679**
Percentage of people who identified themselves as being of Aboriginal or Torres Strait Islander origin (INDIG)	-.488**	-.518**	-.454**	-.469**

** Correlation is statistically significant at p=.01.
+ Omitted from further analyses

Table 2: Variance in aggregated school outcomes explained by the four ABS indices of SES and the 14 SEIFA measures

	Separate scales		Common scale	
	Primary	**Secondary**	**Primary**	**Secondary**
Indicator of Relative Socio-economic Disadvantage (IRSED)	44.8	46.3		
Indicator of Relative Socio-economic Advantage and Disadvantage (IRSAD)	46.3	43.7		
Index of Education and Occupation (IEO)	47.8	45.4		
Index of Economic Resources (IER)	37.1	41.5		
SEIFA 14 variables – All households	62.1	56.2	61.9	54.2
SEIFA 14 variables – Households with school-age children	63.3	57.3	63.1	56.0

Including 'remoteness' and 'percentage of Aboriginal enrolments'

There is evidence in the research literature to support the contention that remoteness and indigeneity also impact on school outcomes. In Australia, socio-economic measures are very highly correlated with remoteness and indigeneity. Consequently much of the impact of these two factors can be assumed to be accounted for by the 14 SEIFA variables. However, it was considered appropriate to determine whether they had any additional influence over and above that associated with the 14 SEIFA variables.

To obtain a measure of remoteness, the Accessibility Remoteness Index of Australia (ARIA) was used to generate average values for the CCDs in which the schools were located.

To obtain an overall measure of indigineity, the percentage of Indigenous enrolments was computed for each school. While the percentage of people who identified themselves as being of Aboriginal or Torres Strait Islander origin was included as one of the 14 SEIFA CCD variables, it was considered important to investigate whether the overall percentage of school enrolments identifying as Indigenous had any additional explanatory power over and above the community measure already included.

The results of the regression analyses are summarised in Table 3.

Table 3: Additional explanatory power obtained by including a measure of remoteness and 'percentage of Indigenous enrolments'

	14 SEIFA variables	**14 SEIFA variables + ARIA + % of Indigenous enrolments**
Primary	61.9	67.9
Secondary	54.2	59.0

Table 3 indicates that the explanatory power of ICSEA would be enhanced by including both the measure of remoteness and the measure of the percentage of Indigenous enrolments.

On the basis of the above analyses, it was decided that:

1. ICSEA would be constructed from the 14 SEIFA variables, school remoteness and the percentage of Indigenous enrolments.

2. The same set of weights would be used for both primary and secondary schools.

3. The index would be computed using CCD values for all households within the school, rather than those of the tested students.

4. In line with ABS SES indices, ICSEA would be scaled to a mean of 1000 and standard deviation of 100 for all Australian schools.

The final equations used to construct the ICSEA values were:

INDEX = 2.125 + INC_LOW_Px1.849 + INC_HIGH_Px-1.053 + CERT_Px-5.501 + DIP_Px1.430 + NOQUAL_Px-5.806 + NOSCHOOL_Px6.50 + NOY12_Px2.934 + UNEMP_Px 0.0 + OCC1_Px2.933 + OCC4_Px0.844 + OCC5_Px0.946 + INDIG_Px-4.367 + ONEPAR_Px-5.364 + NONET_Px-0.965

and, after INDEX had been re-scaled to a mean of 1000 and a standard deviation of 100:

ICSEA = 3.970 + INDEXx-0.021 + $INDEX^2$x2.331e-005 + $INDEX^3$x-6.501e-009 + ARIAx0.020 + pINDIGx-0.023

2. Using ICSEA for the *My School* website

Calculating school ICSEA values

The following steps were taken to calculate an ICSEA value for each school on the *My School* website:

1. Residential addresses for each student in Australia were gathered (without student names or 'de-identified') as well as data about each school's:
   - proportion of Indigenous students;
   - remoteness (based on an agreed Australian Bureau of Statistics (ABS) system which identifies localities on a scale from metropolitan through to provincial, remote, and very remote.)

   For government schools, information was gathered from State and Territory education departments. For non-government schools, information was gathered from the Commonwealth Department of Education, Employment and Workplace Relations.

2. Each address was matched to its ABS Census Collection District (CCD). A CCD is a geographical area (smaller than a postcode or suburb) which contains on average about 220 households. Student residential addresses were matched to their CCD by a process known as geo-coding, where the address is located to a point on the earth's surface (a specific latitude and longitude). The CCD in which the address is located was then identified and the address was linked with the CCD's unique identifying number.

3. The relevant socio-economic status (SES) characteristics of the CCD in which each student at a school lives (known from ABS household census data) were aggregated to the school level. For example, for a school with 100 students, who lived in five different CCDs, the relevant variables from SES data for each of those CCDs would be aggregated, proportionally, up to the school level, so that the average for each SES data variable at the school level were known.

4. The consolidated school level SES data were fed into the ICSEA formula developed in the construction and modelling process described earlier in the paper. The known data about the remoteness of a school and the proportion of Indigenous students at the school were also fed into the formula.

5. An ICSEA value was calculated.

As indicated earlier, ICSEA places schools on a numerical scale by reference to their relative socio-educational advantage. It is scaled to a mean of 1000 and standard deviation of 100 for all Australian schools. For example:

- a school in a regional town with a student population drawn largely from relatively disadvantaged households might have an ICSEA value of about 850;
- a school in a metropolitan area which draws its students from relatively advantaged households might have an ICSEA value of about 1150; and
- a school in a remote Indigenous community might have an ICSEA value of about 540.

Limitations of ICSEA

ICSEA makes use of the same fundamental approach that the Commonwealth has long used to allocate funds to non-government schools, namely to use CCD information on each student as a means of generating an index that best captures contextual characteristics of the school. It differs in that variables were selected that maximise its capacity to predict performance on NAPLAN tests. In other words, it is not a measure of socio-economic status per se, but rather of the socio-educational character of the students within a school.

However, as in the case of the Commonwealth index, it has the limitation that in a small proportion of cases, ICSEA may provide an inappropriate measure of the socio-educational level of the school. This can occur in instances where there is a mismatch between students' actual levels and that of the CCD values associated with their addresses. An example would be remote schools where the ICSEA values are inflated by the presence in CCDs of farmers who send their children to city boarding schools.

To address this limitation, ICSEA values for a small proportion of schools were adjusted for the *My School* website where additional evidence was available to indicate that the CCD values do not properly reflect the student demographics of the school. An expert panel was convened to review the changes and ensure consistency in the criteria used to make them. There will continue to be a need for a formal review process to make ongoing adjustments where there is evidence that ICSEA does not properly reflect the actual circumstances of students in a given school.

Grouping similar schools using ICSEA values

The *My School* website uses school ICSEA values to group and compare schools that serve similar student populations. Schools will be compared with up to 60 other statistically similar schools (indicated by having similar ICSEA values).

Most schools will have 60 statistically similar schools comprising the 30 closest to themselves both above and below them on the ICSEA scale. Some schools will have fewer than 60 schools in their cohort if they are at either the top or bottom end of the ICSEA scale as they may not have 30 schools above or below them.

Comparing schools within statistically similar groups allows the identification of schools which are performing well compared to schools that serve similar student populations. This provides the opportunity to identify and share with other schools the practices within a school's community that are making a difference and improving student outcomes.

Derivation of SES quarters from ICSEA values

As well as being used to group and compare schools that serve similar student populations, the variables that are used to calculate the ICSEA value are also used to derive information about the student population background, shown on the *My School* profile page for a school. The information is shown in the form of quarters. It shows the proportion of educationally advantaged or disadvantaged students in the school compared with the spread of students across Australia. For example, if a school was exactly representative of the range of students across Australia, the quarter percentages would all be 25%.

If a school was providing for students from predominantly advantaged backgrounds, then the top quarter might have 55%, the middle quarters 34% and 11%, and 0% in the bottom quarter.

Conclusion

ICSEA was developed to provide a valid way of grouping schools for the purpose of comparing NAPLAN literacy and numeracy results for students in schools that are statistically similar. It provides a measure that is highly correlated with performance as measured by the NAPLAN tests.