ICSEA 2024

Technical Report

Assessment and Reporting
Measurement and Evaluation Unit

February 2025

Acknowledgement of Country

ACARA acknowledges the Traditional Owners and Custodians of Country and Place throughout Australia and their continuing connection to land, waters, sky and community. We pay our respects to them and their cultures, and Elders past and present.

Copyright

Contact details

Australian Curriculum, Assessment and Reporting Authority
Level 13, Tower B, Centennial Plaza, 280 Elizabeth Street Sydney NSW 2000
T 1300 895 563 | F 1800 982 118 | www.acara.edu.au

Table of contents

Introduction

Guide to ICSEA technical reports

Technical reports relating to the Index of Community Socio-educational Advantage (ICSEA) have been published since 2013.

The *ICSEA 2013: Technical Report* provides a thorough explanation of the methodology underlying ICSEA calculations.

Technical reports from 2014 to 2018 provide an overview of the procedures employed, referring to the *ICSEA 2013: Technical Report* as the primary explanatory document. Outcomes are presented either in the body of the report or as appendices.

In 2019, the similar-students analysis (SSA) was introduced. This model shares much of the same data and methodology as ICSEA. It is used to create a predicted score for each school's NAPLAN results, based on the results of students with a similar socio-educational background. The difference between the predicted and actual scores is used to classify school performance on each test in My School. This enhancement of ICSEA was elaborated in the document *Technical Report 2019: Approach to reporting on My School*.

The *ICSEA 2021: Technical Report* again provided an overview of ICSEA procedures, with the outcomes of the generalised partial credit model and multi-level regression coefficients presented as appendices. The SSA is not described.

The *ICSEA 2022: Technical Report* provided a summary of the calculations underpinning both ICSEA and SSA, explaining the relationships and dependencies between their methodologies, the specifications of all models, and some small refinements that were introduced in the 2022 cycle.

No changes have been introduced to the methodology since 2022. This technical report reiterates the explanation of the process as presented in the 2022 technical report, and presents the 2024 results. Should further technical details be required, please refer to previous technical reports: primarily the 2013 and 2019 editions.

Overview of methodology

Both ICSEA and SSA are calculated by a 3-stage process:

- Item calibration: student background data items are calibrated to construct the SEA scale.
- Conditioning model: plausible values for each student's SEA are drawn, anchoring on parameters that emerged from the previous stage.
- Multi-level model: NAPLAN performance is predicted from SEA and background data.

The item calibration is common between ICSEA and SSA, and is conducted for parental background item responses obtained from the NAPLAN data set (students in Years 3, 5, 7 and 9).

Socio-educational advantage (SEA) plausible values for SSA are drawn for the set of students who complete NAPLAN in that calendar year: students in Years 3, 5, 7 and 9. The multi-level model for SSA uses this set of plausible values for both student SEA and school SEA.

SEA plausible values for ICSEA are drawn for all students in all schools. The multi-level model for ICSEA uses this set of plausible values for school SEA, while the student SEA is taken from the plausible values that were drawn for the NAPLAN data set.

This process is illustrated diagrammatically in Figure 1.

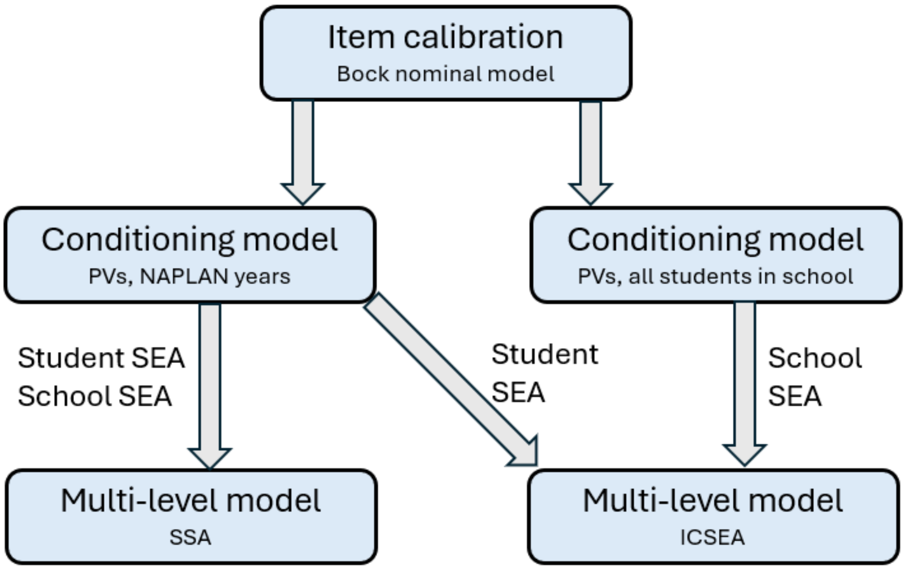

Figure 1: SSA and ICSEA process overview

Item calibration

SEA items

There are 8 student background data items: 4 for each parent. These 4 items are listed, along with their allowable responses.

1. School education (se)
    a. Year 9 or equivalent
    b. Year 10 or equivalent
    c. Year 11 or equivalent
    d. Year 12 or equivalent
2. Non-school education (nse)
    a. No non-school education
    b. Certificate I–IV, including trade certificate
    c. Advanced diploma or diploma
    d. Bachelor degree or above
3. Occupation group (occ)
    a. Unskilled manual, office and sales
    b. Skilled trades, clerical and sales
    c. Other managers and associate professionals
    d. Senior managers and professionals

4. Non-paid work (npo)

    a. In non-paid occupation

    b. In paid occupation

Of the 8 items used to construct the SEA scale, 6 are partial credit items with a maximum score of 3, and 2 are dichotomous items (1/0).

Methodology

Data from all jurisdictions and all NAPLAN year levels was included and senate weights were applied, ensuring each jurisdiction contributed equally to the item calibration process.

Item response theory (IRT) was used to calibrate the student background data items.

- The Bock nominal model (Bock 1972) was applied. This allows the difference between item scores to vary between response levels, so that, for instance, the difference between scores for (school education = Year 12 or equivalent) and (school education = Year 11 or equivalent) is no longer constrained to be equal to the difference between scores for (school education = Year 11 or equivalent) and (school education = Year 10 or equivalent).
- The scores for parent 1 and parent 2 were constrained to be equal, for all items. This overcomes the difficulty of specifying which parent should be designated as parent 1 and which as parent 2.

The item calibration process is shown diagrammatically in Figure 2. In this and other figures, the outputs shaded purple are those that are published in the technical report.

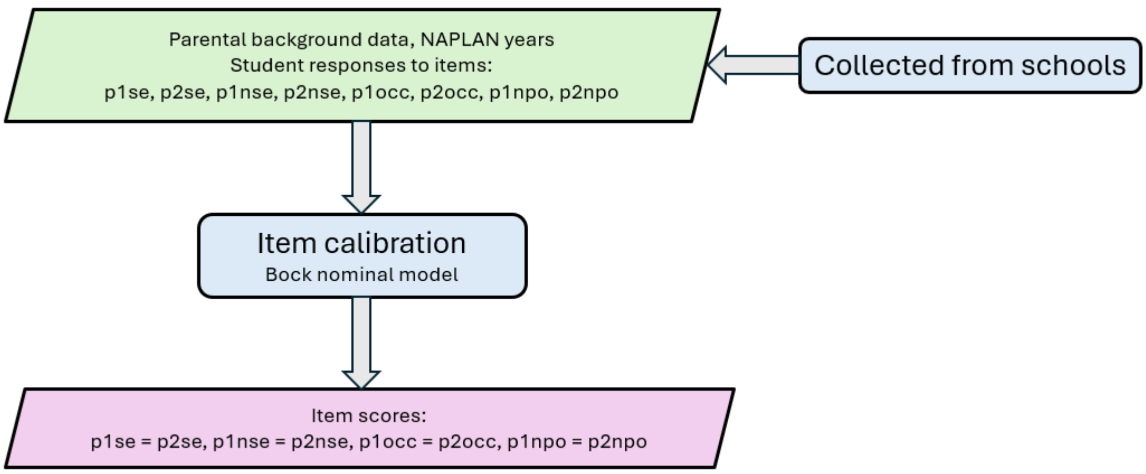

Figure 2: Item calibration

Results

Table 1 to Table 8 show the item calibration results for each of the 8 items, alongside their calibration results in 2023.

The “Response” column shows the response category available to the parental question. The “Count” column shows the number of instances of a particular response. The “%” column shows the percentage that the number of instances amounted to. The “Code” column provides the ordered coded response categories. The “2024” and “2023” columns show the estimated item scores obtained from the calibration model for the corresponding ICSEA calculation cycle.

Two points should be noted:

- The estimated scores are not equidistant between adjacent response levels. This flexibility is a result of the introduction of the Bock nominal model, from 2022.
- Items relating to parent 1 and parent 2 have been constrained to have identical scores. Table 1 and Table 2 have the same estimated item scores, as do Table 3 and

- Table 4, Table 5 and Table 6, and
- Table 7 and Table 8.

Table 1: Parent 1: school education (p1se)

Response	Count	%	Code	2024	2023
Year 9 or equivalent	59988	4.98	0	0.00	0.00
Year 10 or equivalent	167504	13.92	1	1.45	1.52
Year 11 or equivalent	109937	9.14	2	1.73	1.83
Year 12 or equivalent	865989	71.96	3	3.61	3.68

Table 2: Parent 2: school education (p2se)

Response	Count	%	Code	2024	2023
Year 9 or equivalent	50743	4.92	0	0.00	0.00
Year 10 or equivalent	161314	15.63	1	1.45	1.52
Year 11 or equivalent	93735	9.08	2	1.73	1.83
Year 12 or equivalent	726571	70.38	3	3.61	3.68

Table 3: Parent 1: non-school education (p1nse)

Response	Count	%	Code	2024	2023
No non-school education	214285	18.21	0	0.00	0.00
Certificate I–IV inc. trade certificate	314799	26.74	1	1.39	1.43
Advanced diploma or diploma	165026	14.02	2	2.83	2.91
Bachelor degree or above	482947	41.03	3	4.72	4.83

Table 4: Parent 2: non-school education (p2nse)

Response	Count	%	Code	2024	2023
No non-school education	165019	16.48	0	0.00	0.00
Certificate I–IV inc. trade certificate	313243	31.28	1	1.39	1.43
Advanced diploma or diploma	135768	13.56	2	2.83	2.91
Bachelor degree or above	387234	38.67	3	4.72	4.83

Table 5: Parent 1: occupation (p1occ)

Response	Count	%	Code	2024	2023
Unskilled manual, office and sales	165332	17.31	0	0.00	0.00
Skilled trades, clerical and sales	250984	26.28	1	0.88	0.91
Other managers and associate professionals	229291	24.01	2	2.02	2.09
Senior managers and professionals	309314	32.39	3	3.27	3.39

Table 6: Parent 2: occupation (p2occ)

Response	Count	%	Code	2024	2023
Unskilled manual, office and sales	173115	19.38	0	0.00	0.00
Skilled trades, clerical and sales	244245	27.34	1	0.88	0.91
Other managers and associate professionals	215813	24.16	2	2.02	2.09
Senior managers and professionals	260063	29.11	3	3.27	3.39

Table 7: Parent 1: non-paid occupation (p1npo)

Response	Count	%	Code	2024	2023
In non-paid occupation	200660	17.36	0	0.00	0.00
In paid occupation	954921	82.64	1	1.08	1.11

Table 8: Parent 2: non-paid occupation (p2npo)

Response	Count	%	Code	2024	2023
In non-paid occupation	105900	10.60	0	0.00	0.00
In paid occupation	893236	89.40	1	1.08	1.11

Conditioning model: NAPLAN years

Methodology

Plausible values for socio-educational advantage (SEA) were drawn for all students in the NAPLAN years (consisting only of students in Years 3, 5, 7 and 9).

The set of items is as described in the section “SEA items”.

The following conditioning variables were used:

- wler: NAPLAN reading weighted likelihood estimate
- mwler: a dummy variable indicating whether wler is missing
- g1-g4: school geolocation, where g1 = inner regional, g2 = outer regional, g3 = remote and g4 = very remote. The reference category is g0 (major cities)
- atsi Aboriginal and Torres Strait Islander status
- matsi: a dummy variable indicating whether atsi is missing.

The conditioning model allows SEA plausible values to be drawn for students who have missing responses to some or all of the parental background data items.

The resulting 5 SEA plausible values for each NAPLAN student are used as student-level SEA estimates in the multi-level modelling that provides the final SSA regression equation, along with the residuals from that regression, which are used to categorise school performance.

For the school-level SEA estimates used in the SSA multi-level modelling, the student-level SEA estimates are averaged across all students in each year level at the school.

The process by which plausible values are drawn for students in the NAPLAN years is shown diagrammatically in Figure 3.

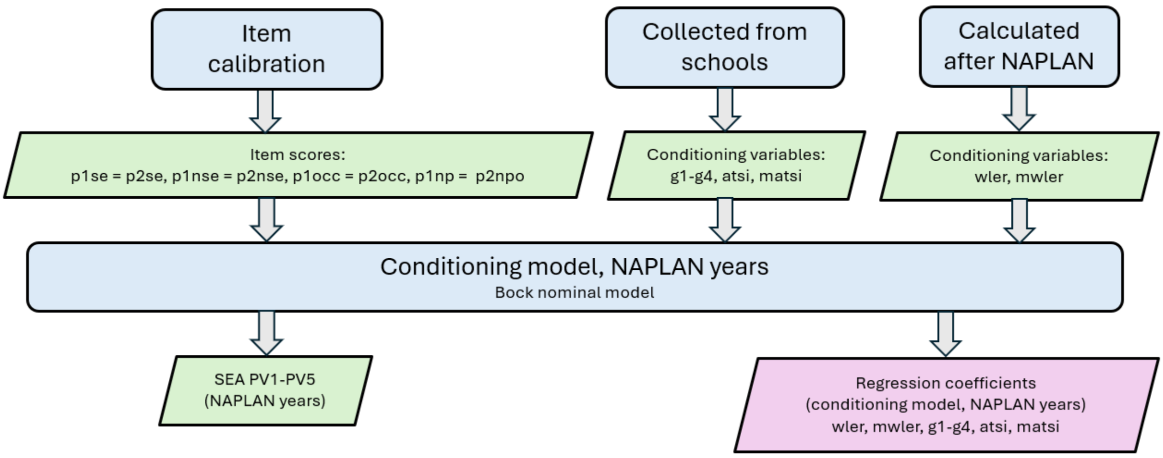

Figure 3: Conditioning model – NAPLAN years

Results

Regression coefficients from the conditioning model are shown in Table 9. While these do not serve as inputs to any further modelling, they are published here for reproducibility of the analysis.

Table 9: Regression coefficients – conditioning model, NAPLAN data set

Regression variable	Coefficient
(constant)	0.204
wler	0.409
mwler	−0.531
g1	−0.292
g2	−0.351
g3	−0.320
g4	−0.487
atsi	−0.599
matsi	0.165

Conditioning model: all students in school

Methodology

Plausible values for SEA were drawn for all students at each school.

The set of items is as described in the section "SEA items".

The following conditioning variables were used:

- schwler: the school average of NAPLAN reading weighted likelihood estimates
- g1-g4: school geolocation, where g1 = inner regional, g2 = outer regional, g3 = remote and g4 = very remote. The reference category is g0 (major cities)
- atsi: Aboriginal and Torres Strait Islander status
- matsi: a dummy variable indicating whether ATSI is missing.

The conditioning model allows SEA plausible values to be drawn for students who have missing responses to some or all of the parental background data items.

For the school-level SEA estimates used in the ICSEA multi-level model, the student-level SEA estimates are averaged across all students in the school.

These student-level SEA plausible values are also used to define the "SEA quarters" that are used in NAPLAN national results and published on the My School website.

However, the student-level SEA estimates used in the ICSEA multi-level model are taken from the conditioning model run for students in the NAPLAN years.

The process by which plausible values are drawn for all students in each school is shown diagrammatically in Figure 4.

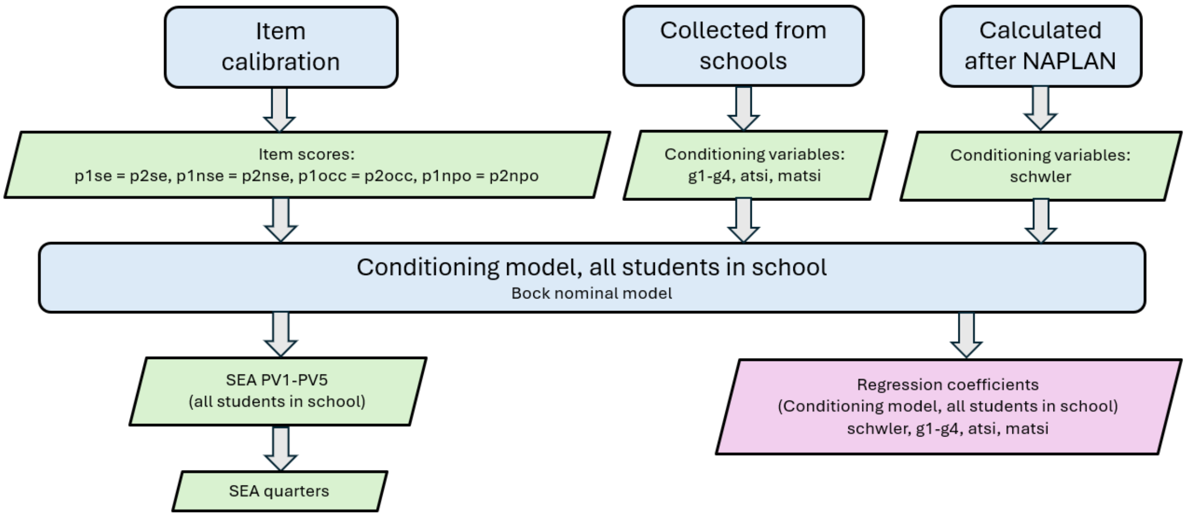

Figure 4: Conditioning model – all students in school

Results

Regression coefficients from this conditioning model are shown in Table 10. While these do not serve as inputs to any further modelling, they are published here for reproducibility.

Table 10: Regression coefficients – conditioning model, all students in school

Regression variable	Coefficient
(constant)	0.135
schwler	1.292
g1	−0.099
g2	−0.047
g3	0.117
g4	0.091
atsi	−0.541
matsi	0.016

Multi-level model: similar-students analysis

Methodology

The similar-students analysis can be regarded as comparing a school's average achievement with the average achievement of Australian students with a similar background to the students in that school. The average achievement of students with a similar background is determined as the predicted score from a multi-level regression model; the difference is expressed as the residual for each school.

A multi-level regression model (MLM) with a school and a student level was applied to predict NAPLAN scores from the inputs of socio-educational advantage (SEA), Aboriginal and Torres Strait Islander status (ATSI), and remoteness of the school, which is measured by the Accessibility/Remoteness Index of Australia (ARIA).

The system of equations for the random intercept, fixed slopes MLM is:

Level 1 (student):

$$Y_{ij} = \beta_{0j} + \beta_1 SEA_{ij} + \beta_2 ATSI_{ij} + \beta_3 MATSI_{ij} + r_{ij}$$

Level 2 (school):

$$\beta_{0j} = \gamma_{00} + \gamma_{01} SEA_j + \gamma_{02} ATSI_j + \gamma_{03} ARIA_j + \mu_{0j}$$

Each term is defined as follows, with SEA scores obtained from the plausible values drawn for the NAPLAN data set:

Y_{ij}	the performance of student *i* in school *j* for that domain
SEA_{ij}	the SEA score for student *i* in school *j*
$ATSI_{ij}$	the Aboriginal and Torres Strait Islander status of student *i* in school *j*
$MATSI_{ij}$	an indicator of whether the Aboriginal and Torres Strait Islander status information is missing
r_{ij}	the residual for student *i* in school *j*
SEA_j	the mean SEA score of students in the same year level of school *j*
$ATSI_j$	the percentage of Aboriginal and Torres Strait Islander students in school *j*
$ARIA_j$	the Accessibility/Remoteness Index of Australia, which is a measure of the remoteness of school *j*
μ_{0j}	the residual for school *j*

Combining the equations gives the following full model:

$$Y_{ij} = \gamma_{00} + \gamma_{01} SEA_j + \gamma_{02} ATSI_j + \gamma_{03} ARIA_j + \beta_1 SEA_{ij} + \beta_2 ATSI_{ij} + \beta_3 MATSI_{ij} + \mu_{0j} + r_{ij}$$

The school-level residuals (μ_{0j}) underpin the reporting of school performance against similar schools on My School. Positive residuals indicate higher achievement than predicted; negative residuals indicate lower achievement than predicted.

Further details can be found in the *Technical Report 2019: Approach to reporting on My School*. These details include calculation of standard errors, exclusions from the data set and explanation of presentation on My School.

The data flow for the SSA multi-level model is shown diagrammatically in Figure 5.

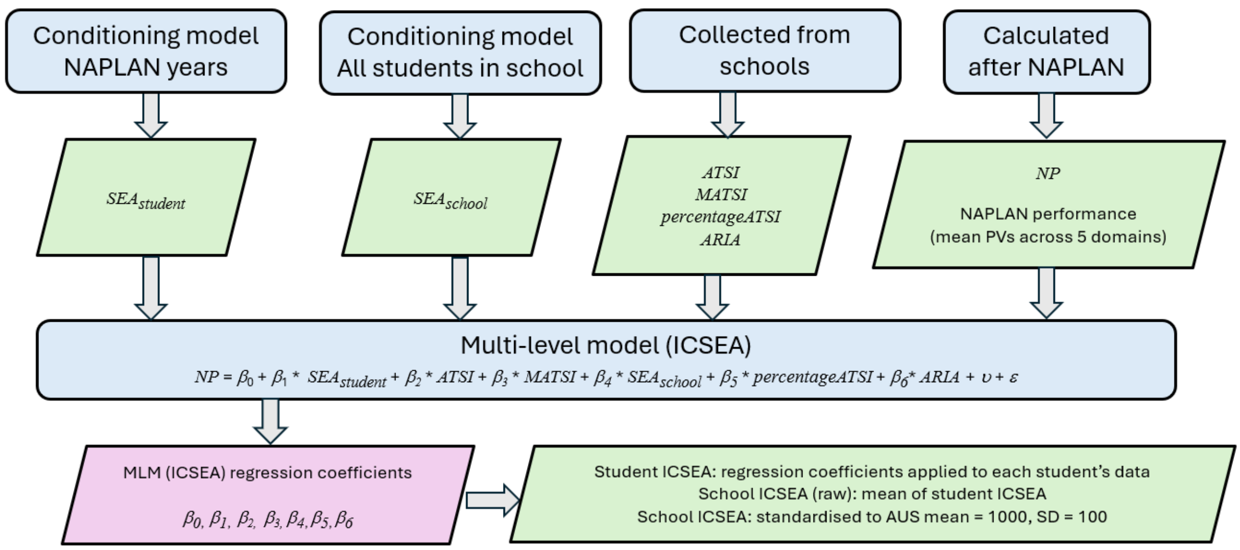

Figure 5: Multi-level model: SSA

Results

The regression coefficients obtained from the multi-level model are shown in Table 11. Note that the regression model is run separately for each year level and domain.

Table 11: Regression coefficients – multi-level model: SSA

Domain	Year	γ_{00}	γ_{01}	γ_{02}	γ_{03}	β_1	β_2	β_3
		Intercept	SEA school	ATSI school	ARIA school	SEA student	ATSI student	MATSI student
N	3	405.966	21.196	−0.440	−0.100	19.692	−18.611	−12.365
N	5	491.345	24.280	−0.451	−0.273	21.754	−23.640	−11.832
N	7	541.473	39.583	−0.454	0.834	21.434	−27.516	−11.869
N	9	569.915	39.530	−0.474	1.685	20.187	−26.564	−11.712
R	3	406.911	24.104	−0.568	−0.273	27.121	−21.294	−14.859
R	5	496.098	22.164	−0.567	−0.427	24.888	−22.894	−10.947
R	7	538.282	33.211	−0.566	0.754	23.358	−24.224	−11.867
R	9	570.683	31.792	−0.551	1.687	22.944	−23.649	−13.889
W	3	420.613	19.249	−0.647	−0.826	15.846	−22.999	−13.171
W	5	490.527	19.217	−0.721	−1.725	16.543	−22.751	−10.608
W	7	546.222	30.823	−0.894	−0.561	17.392	−26.590	−15.014
W	9	582.343	34.975	−0.938	1.018	19.191	−30.105	−15.965
S	3	405.402	16.824	−0.621	−1.733	20.587	−22.749	−11.308
S	5	490.675	15.730	−0.526	−2.354	18.580	−23.042	−9.351
S	7	543.935	22.069	−0.460	−1.565	16.062	−19.828	−10.056

S	9	571.223	19.874	−0.317	−0.805	13.621	−17.628	−7.725
G	3	411.814	26.971	−0.526	−1.088	27.412	−23.960	−16.227
G	5	502.043	22.295	−0.504	−0.865	23.575	−26.816	−13.429
G	7	540.719	36.134	−0.555	0.323	23.124	−29.079	−13.872
G	9	562.054	38.934	−0.566	1.201	24.446	−29.529	−17.236

Multi-level model: ICSEA

Methodology

The multi-level model for ICSEA has both student and school levels. Its structure is similar to that used for SSA.

$$NP = \beta_0 + \beta_1 * SEA_{student} + \beta_2 * ATSI + \beta_3 * MATSI + \beta_4 * SEA_{school} + \beta_5 * percentageATSI + \beta_6 * ARIA + \upsilon + \varepsilon$$

Each term is defined as follows, with student-level SEA scores obtained from the plausible values drawn for the NAPLAN data set, and school-level SEA scores obtained from the plausible values drawn for all students in the school:

NP	average NAPLAN performance of each student across all domains
$SEA_{student}$	the SEA score for each student
ATSI	the Aboriginal and Torres Strait Islander status of each student
MATSI	an indicator of whether the Aboriginal and Torres Strait Islander status information is missing
SEA_{school}	the mean SEA score of students in the school
$percentageATSI$	the percentage of Aboriginal and Torres Strait Islander students in the school
ARIA	the Accessibility/Remoteness Index of Australia, which is a measure of the remoteness of the school
υ	the school-level residual
ε	the student-level residual

The data flow for the ICSEA multi-level model is shown diagrammatically in Figure 6.

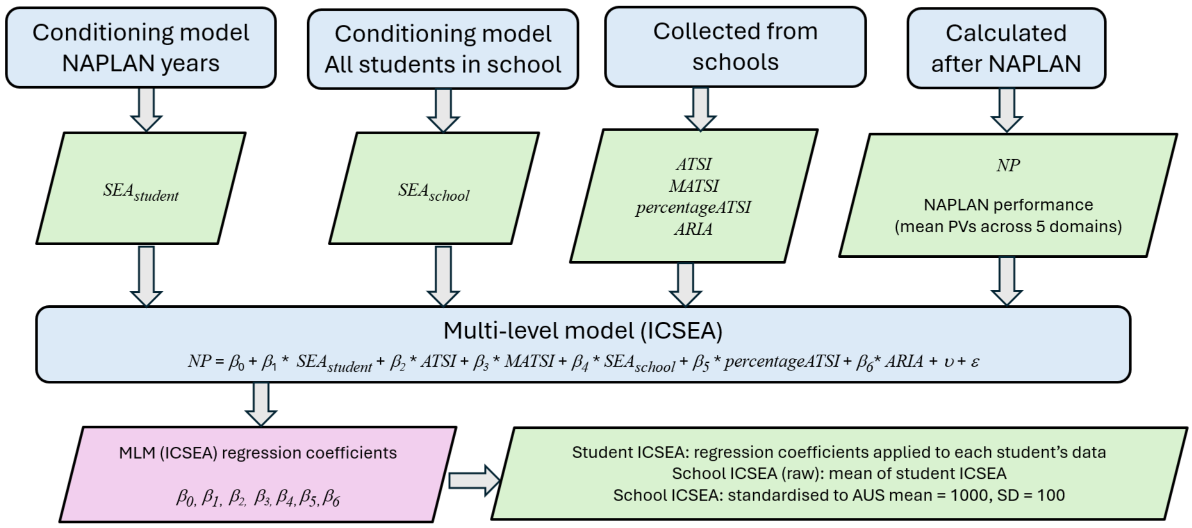

Figure 6: Multi-level model: ICSEA

Results

Regression coefficients

The regression coefficients obtained from the multi-level model are shown in Table 12. The coefficients calculated from all 5 plausible values are shown, and are consistent.

Table 12: Regression coefficients – multi-level model: ICSEA

Coefficient	Variable	PV1	PV2	PV3	PV4	PV5
β_0	(intercept)	0.028	0.026	0.026	0.026	0.027
β_1	$SEA_{student}$	0.249	0.248	0.249	0.249	0.250
β_2	$ATSI$	−0.312	−0.313	−0.313	−0.311	−0.312
β_3	$MATSI$	−0.155	−0.155	−0.158	−0.153	−0.154
β_4	SEA_{school}	0.310	0.313	0.313	0.311	0.311
β_5	$percentageATSI$	−0.006	−0.005	−0.005	−0.005	−0.005
β_6	$ARIA$	−0.011	−0.011	−0.012	−0.011	−0.011

Post-model ICSEA calculations

Once these coefficients have been determined, student-level ICSEA is calculated as follows, for all students in each school:

$$ICSEA_{student} = \widehat{\beta_0} + \widehat{\beta_1} * SEA_{student} + \widehat{\beta_2} * ATSI + \widehat{\beta_3} * MATSI + \widehat{\beta_4} * SEA_{school} + \widehat{\beta_5} * percentageATSI + \widehat{\beta_6} * ARIA$$

The school-level SEA is calculated by averaging student-level SEA estimates for all students in the school; $percentageATSI$ and $ARIA$ are also school-level variables.

All $ICSEA_{student}$ values within a school are then averaged to obtain each school's raw ICSEA.

Raw school ICSEA values are then standardised to a mean of 1000 and a standard deviation of 100.

Stability of ICSEA over time

Figure 7 shows the comparison of the ICSEA in 2023 and 2024. The black line represents a least-squares regression fit. The black cross shows the median in the horizontal and vertical axes. The boxplots at the top and right end of the graph are a representation of each distribution, where the median, the interquartile range, whiskers at 1.5 interquartile range and the individual points considered as outliers (outside the whiskers) are represented for each dimension. These representations are used in all the following figures.

As is shown, the regression line has a slope of 0.99 and explained variance is 98.5%, indicating a very strong positive correlation. Outliers are almost invariably schools with very low enrolments.

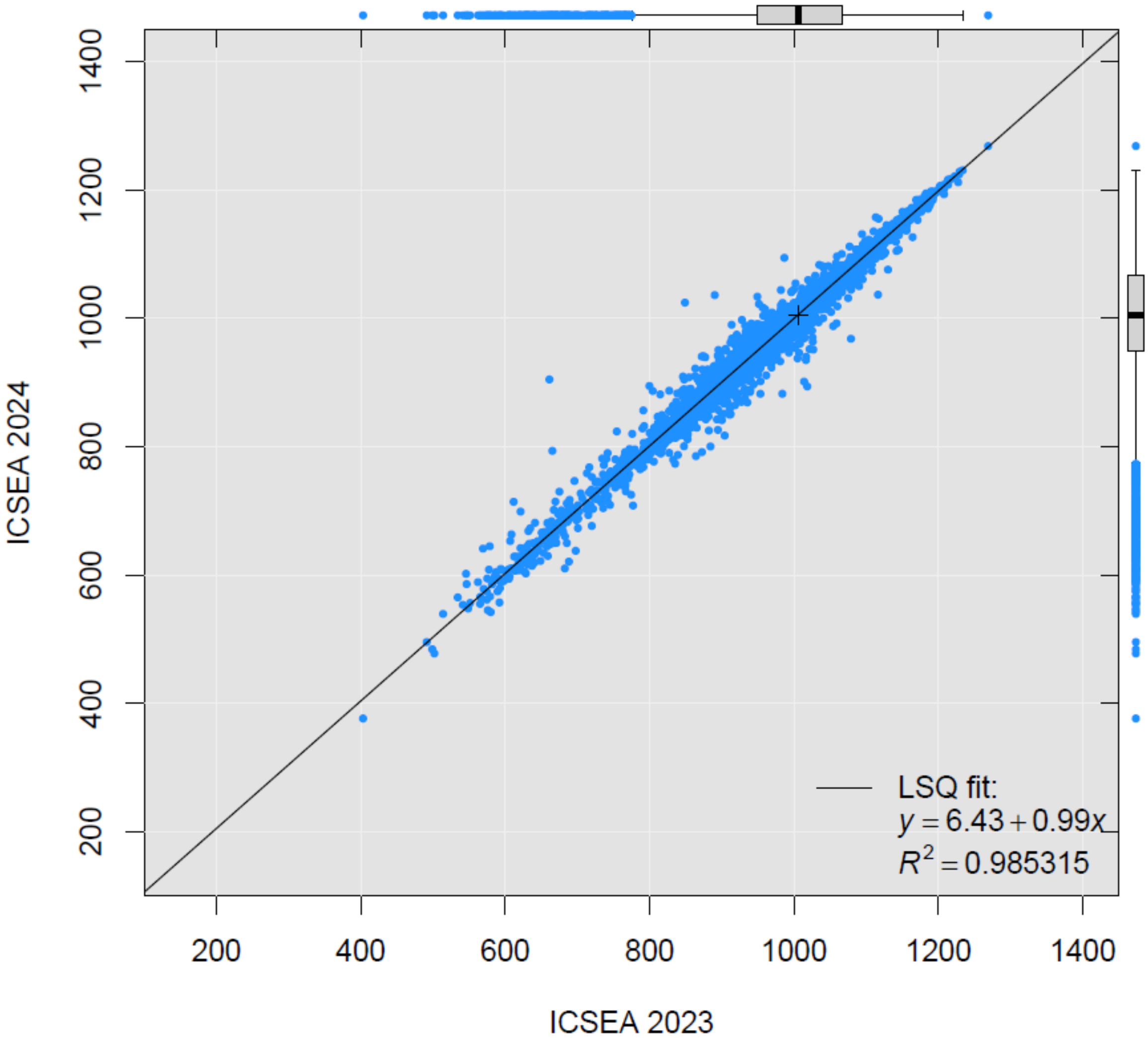

Figure 7: ICSEA 2023 compared with ICSEA 2024

ICSEA as a predictor of NAPLAN performance

Figure 8 shows the scatterplot between published 2024 ICSEA and averaged school performance across all NAPLAN 2023 tests and all year levels available in a school. The regression analysis shows that 72% of variance in school performance is accounted for by ICSEA.

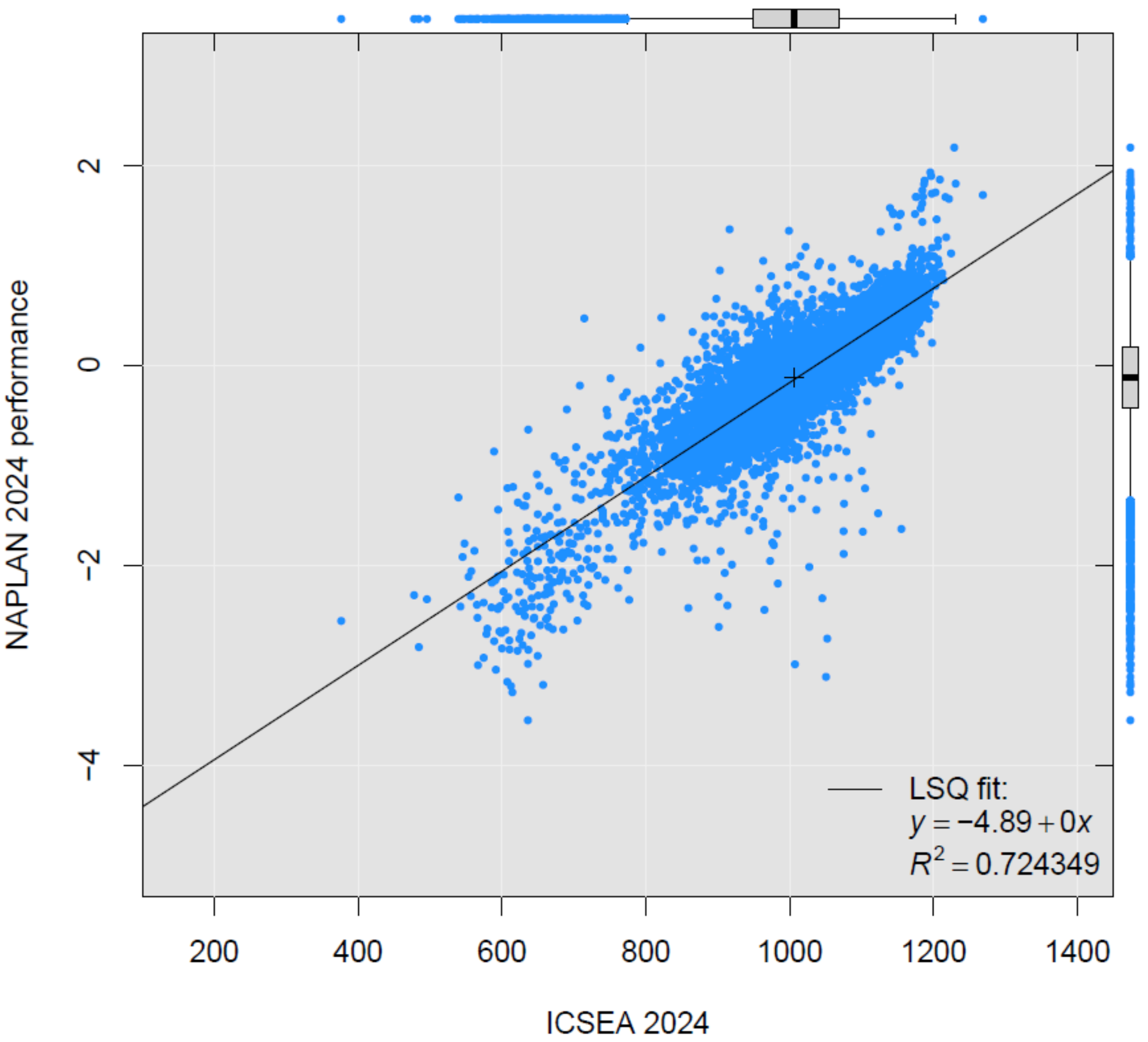

Figure 8: ICSEA 2024 compared with NAPLAN performance

Weighted sum of SEA quarters

The SEA quarters are a broad representation of a school's student distribution. Since 2013, this index has been based solely on each student's level of socio-educational advantage as defined by the student's parental education and occupation – as opposed to ICSEA, which applies a further multi-level modelling step. It is calculated from the SEA plausible values drawn for all students in the school.

For each school, a weighted sum of SEA quarters was calculated as follows:

*Sum SEA quarters = percentage Q1*1 + percentage Q2*2 + percentage Q3*3 + percentage Q4*4*

This weighted sum is one measure of socio-educational advantage. It does not serve the same purpose as ICSEA, but is positively correlated with it.

Figure 9 shows a scatterplot between the weighted sum of SEA quarters and ICSEA in 2023. The relationship is similar to that exhibited in previous years: positively correlated, but not linearly.

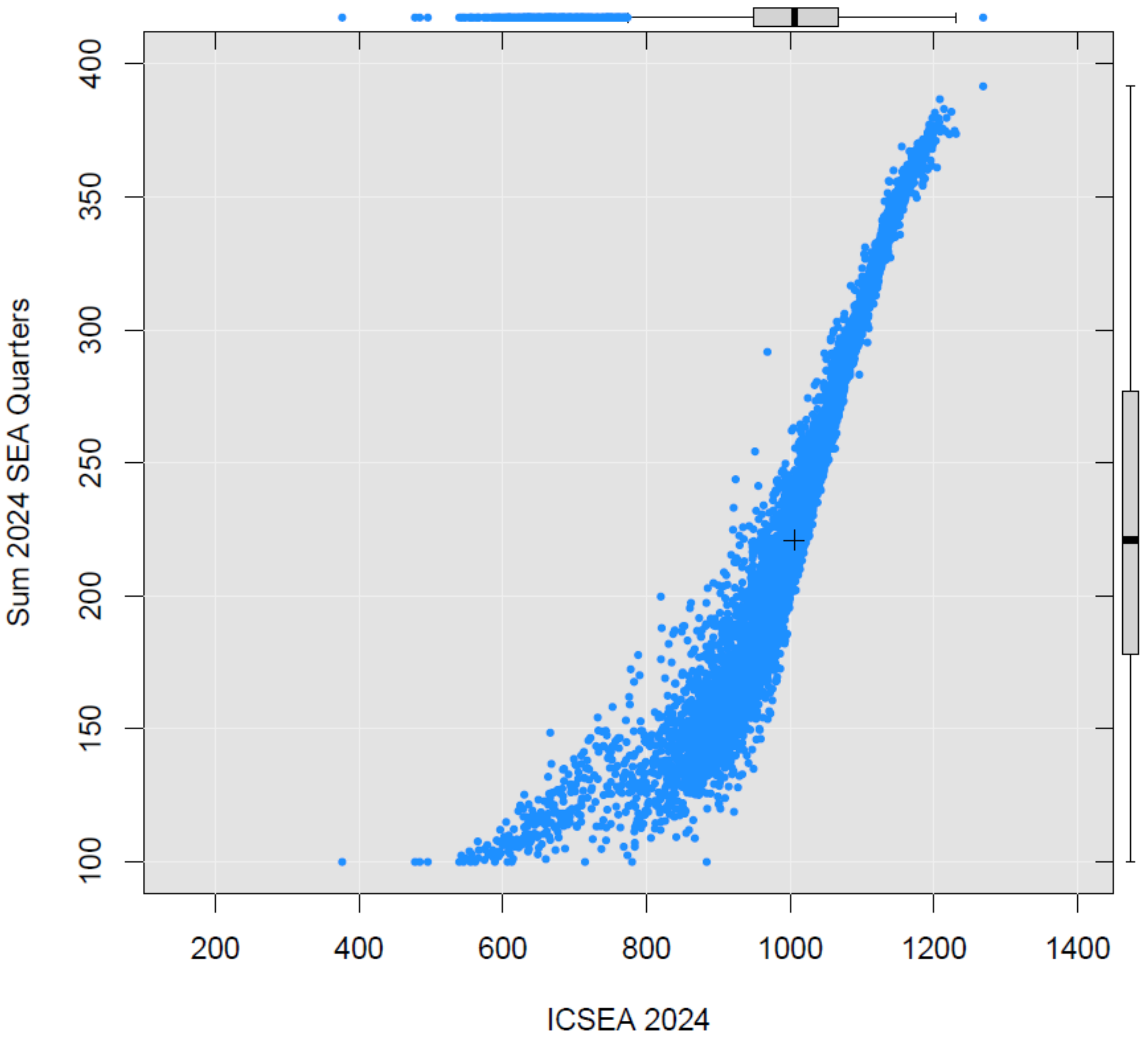

Figure 9: ICSEA 2024 compared with sum of SEA quarters

References

Bock DR (1972) “Estimating item parameters and latent ability when responses are scored in two or more nominal categories”, *Psychometrika*, 37:29–51, https://link.springer.com/article/10.1007/BF02291411, accessed 16 September 2024.