ICSEA 2022

Technical Report

Assessment and Reporting
Measurement and Evaluation Unit

September 2024

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, and 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.

This 2022 technical report provides a summary of the calculations underpinning both ICSEA and SSA. It shows the relationships and dependencies between their methodologies, sets out the specifications of all models, and explains some small refinements that were introduced in the 2022 cycle.

Further detail can be obtained by referring to the 2013 and 2019 technical reports.

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 the school. 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.

The process is illustrated diagrammatically in Figure 1.

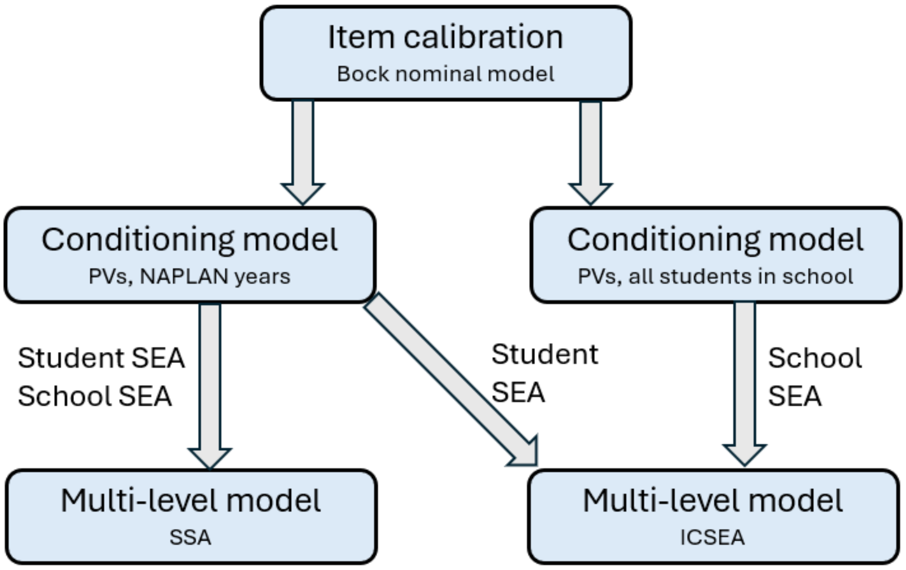

Figure 1: SSA and ICSEA process overview

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. When last described in the *Technical Report 2019: Approach to reporting on My School*, the generalised partial credit model (Adams et. al. 2020) was used to estimate model parameters. However, in 2022 the model was refined in 2 ways:

- 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.

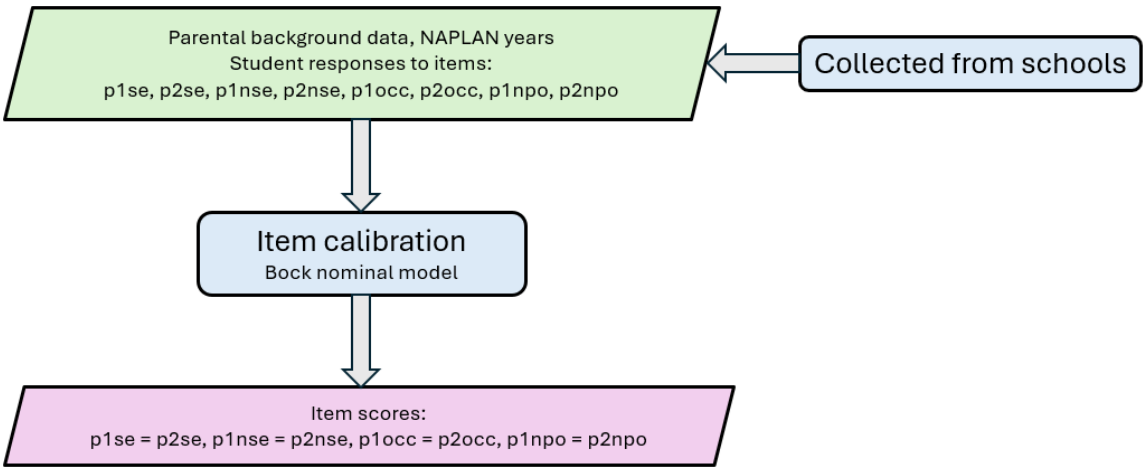

Figure 2: Item calibration

Results

Table 1 to Table 8 show the item calibration results for each of the 8 items, alongside their calibration results in 2021.

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 "2022" and "2021" columns show the estimated item scores obtained from the calibration model for the corresponding ICSEA calculation cycle.

Two points should be noted:

- Firstly, note the estimated scores between 2021 and 2022 in each table. In 2021, these scores were equidistant between adjacent response levels. In 2022, because of the application of the Bock nominal model, they are not.
- Secondly, note that the items relating to parent 1 and parent 2 have been constrained to have identical scores in 2022, whereas in 2021 they varied. Table 1 and Table 2 have the same estimated item scores for 2022, 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	2022	2021
Year 9 or equivalent	55539	4.67	0	0	0
Year 10 or equivalent	163799	13.78	1	1.61	1.11
Year 11 or equivalent	99267	8.35	2	1.89	2.22
Year 12 or equivalent	870455	73.21	3	3.74	3.33

Table 2: Parent 2: school education (p2se)

Response	Count	%	Code	2022	2021
Year 9 or equivalent	50705	4.94	0	0	0
Year 10 or equivalent	160977	15.69	1	1.61	0.98
Year 11 or equivalent	85201	8.31	2	1.89	1.97
Year 12 or equivalent	728822	71.06	3	3.74	2.95

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

Response	Count	%	Code	2022	2021
No non-school education	200343	17.23	0	0	0
Certificate I–IV inc. trade certificate	319542	27.49	1	1.33	1.53
Advanced diploma or diploma	182279	15.68	2	2.78	3.07
Bachelor degree or above	460347	39.6	3	4.73	4.60

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

Response	Count	%	Code	2022	2021
No non-school education	160454	16.11	0	0	0
Certificate I–IV inc. trade certificate	324488	32.59	1	1.33	1.50
Advanced diploma or diploma	138707	13.93	2	2.78	3.01
Bachelor degree or above	372060	37.37	3	4.73	4.51

Table 5: Parent 1: occupation (p1occ)

Response	Count	%	Code	2022	2021
Unskilled manual, office and sales	163259	17.81	0	0	0
Skilled trades, clerical and sales	254058	27.72	1	0.95	1.13
Other managers and associate professionals	229348	25.02	2	2.13	2.27
Senior managers and professionals	269823	29.44	3	3.50	3.40

Table 6: Parent 2: occupation (p2occ)

Response	Count	%	Code	2022	2021
Unskilled manual, office and sales	175179	19.54	0	0	0
Skilled trades, clerical and sales	252685	28.18	1	0.95	1.22
Other managers and associate professionals	222054	24.76	2	2.13	2.44
Senior managers and professionals	246781	27.52	3	3.50	3.66

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

Response	Count	%	Code	2022	2021
In non-paid occupation	230816	20.12	0	0	0
In paid occupation	916488	79.88	1	1.05	0.92

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

Response	Count	%	Code	2022	2021
In non-paid occupation	103865	10.38	0	0	0
In paid occupation	896699	89.62	1	1.05	0.87

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.

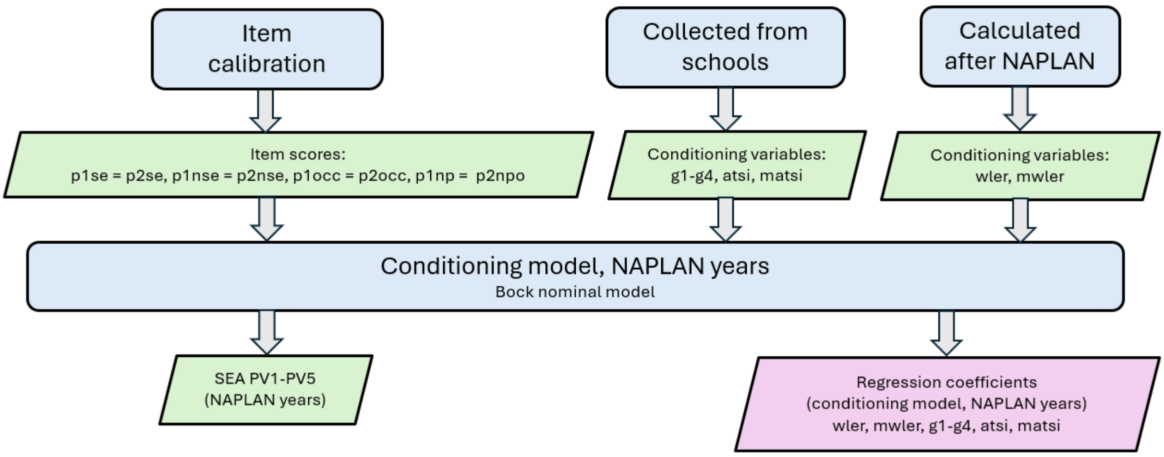

Figure 3: Conditioning model – NAPLAN years

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)	0.192
wler	0.420
mwler	-0.484
g1	-0.285
g2	-0.345
g3	-0.295
g4	-0.483
atsi	-0.600
matsi	0.188

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.

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.

These student-level SEA plausible values are used to define the "SEA quarters" that are used in NAPLAN national results.

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

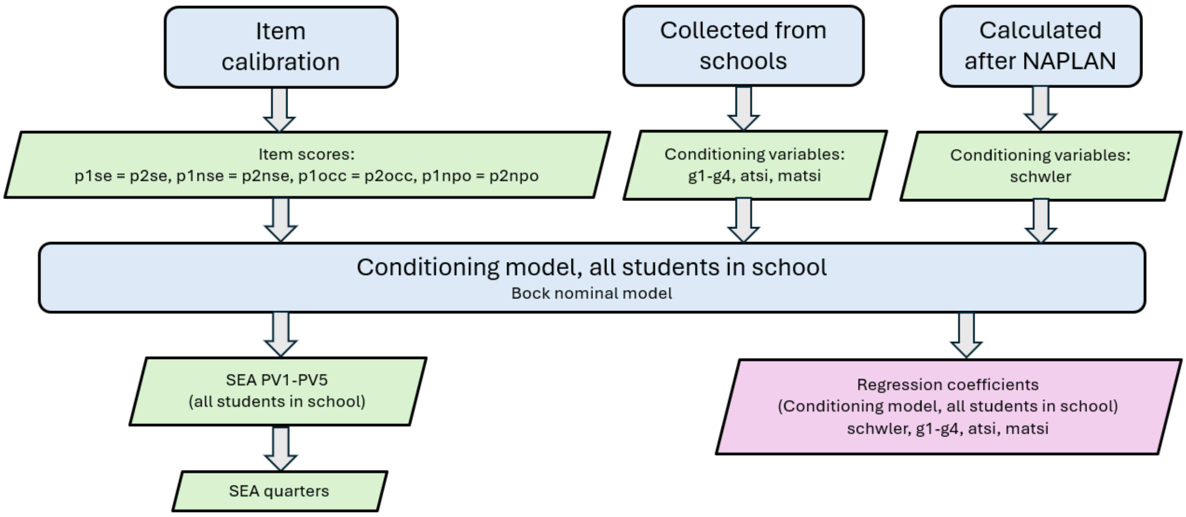

Figure 4: Conditioning model – all students in school

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, SBD data set

Regression variable	Coefficient
(constant)	0.082
schwler	1.350
g1	-0.109
g2	-0.061
g3	0.145
g4	0.156
atsi	-0.530
matsi	0.037

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.

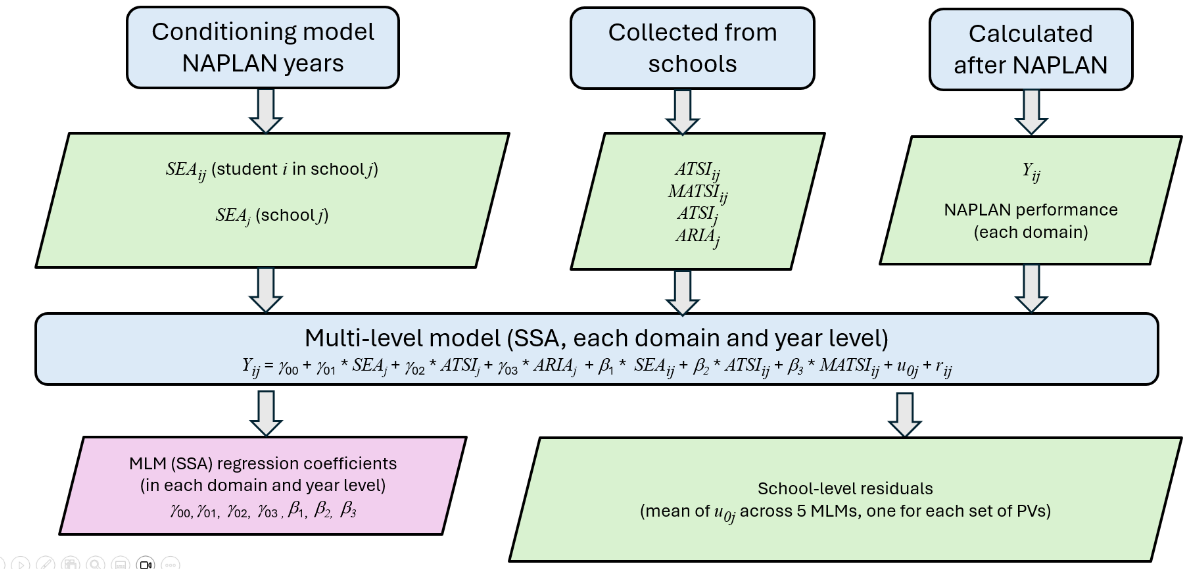

Figure 5: Multi-level model: SSA

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} Intercept	γ_{01} SEA school	γ_{02} ATSI school	γ_{03} ARIA school	β_1 SEA student	β_2 ATSI student	β_3 MATSI student
N	3	400.960	22.143	-0.397	0.031	21.453	-21.158	-16.359
N	5	489.961	19.479	-0.378	0.059	19.113	-20.668	-11.632
N	7	548.832	35.722	-0.421	0.832	21.374	-28.932	-17.208
N	9	587.554	30.673	-0.267	1.191	15.903	-19.935	-14.078
R	3	439.180	26.359	-0.468	-0.429	28.371	-22.118	-18.098
R	5	513.415	17.275	-0.559	-0.040	21.620	-21.328	-12.209
R	7	545.441	25.842	-0.421	0.717	19.622	-21.272	-12.171
R	9	583.039	27.789	-0.478	1.234	20.338	-21.468	-13.443
W	3	425.809	18.231	-0.573	-0.655	14.739	-23.221	-15.712
W	5	489.740	17.990	-0.699	-1.501	16.074	-23.100	-12.468
W	7	535.525	27.561	-0.696	-0.986	16.126	-25.275	-14.303
W	9	567.336	30.506	-0.724	0.339	17.547	-26.182	-14.834
S	3	420.769	20.164	-0.557	-1.892	21.287	-27.390	-16.389
S	5	507.767	15.140	-0.431	-1.850	17.146	-18.783	-8.978
S	7	550.835	23.259	-0.333	-1.333	16.384	-21.668	-11.423

S	9	581.159	21.722	-0.269	-0.689	14.000	-15.631	-8.950
G	3	434.515	27.201	-0.464	-0.580	27.865	-27.000	-20.467
G	5	501.035	18.718	-0.401	-0.496	21.121	-21.062	-13.614
G	7	536.871	28.616	-0.385	0.232	19.950	-27.499	-16.909
G	9	579.672	31.972	-0.414	0.647	21.573	-26.668	-15.018

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.

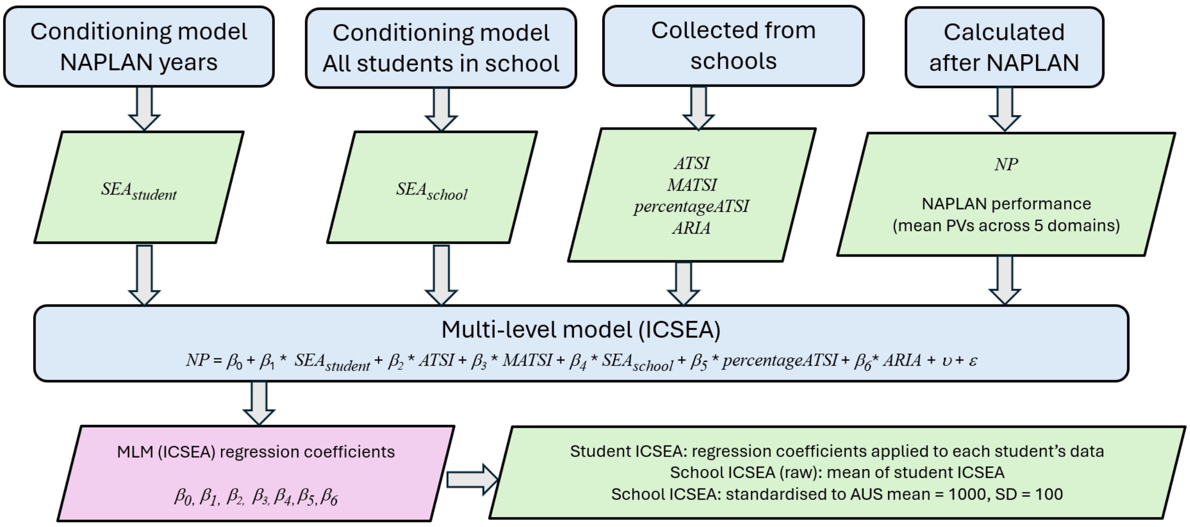

Figure 6: Multi-level model: ICSEA

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)	0.049	0.049	0.050	0.051	0.048
β_1	$SEA_{student}$	0.196	0.195	0.196	0.195	0.195
β_2	$ATSI$	-0.421	-0.423	-0.421	-0.419	-0.422
β_3	$MATSI$	-0.208	-0.208	-0.207	-0.202	-0.208
β_4	SEA_{school}	0.375	0.376	0.376	0.375	0.378
β_5	$percentageATSI$	-0.006	-0.006	-0.006	-0.006	-0.005
β_6	$ARIA$	-0.014	-0.015	-0.014	-0.015	-0.014

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 2021 and 2022. 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%, indicating a very strong positive correlation. Outliers are almost invariably schools with very low enrolments.

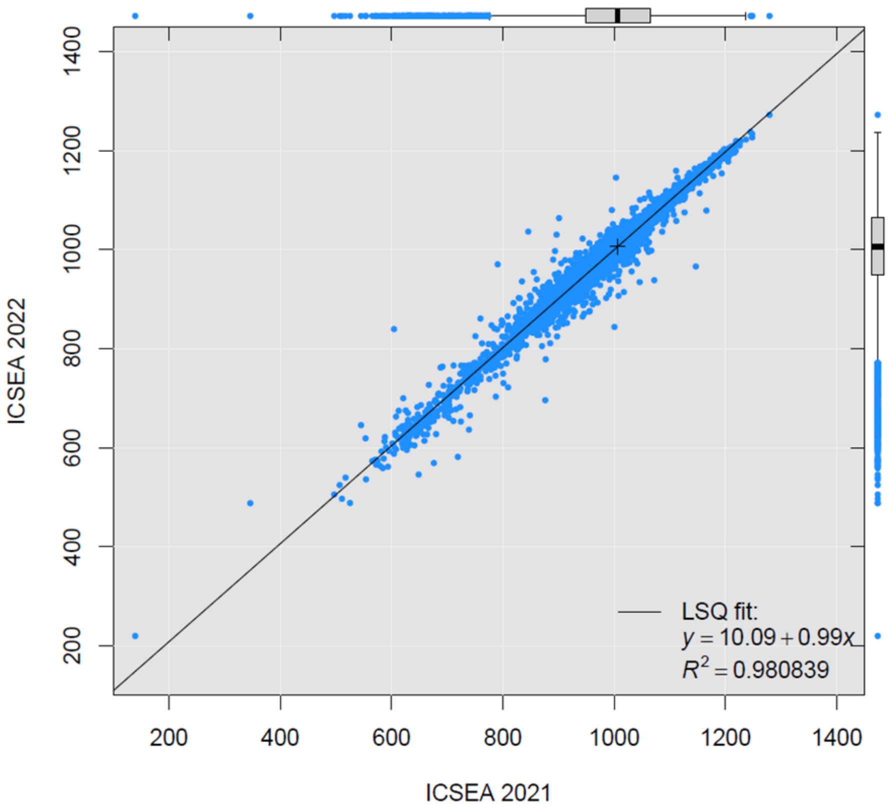

Figure 7: ICSEA 2021 compared with ICSEA 2022

ICSEA as a predictor of NAPLAN performance

Figure 8 shows the scatterplot between published 2022 ICSEA and averaged school performance across all NAPLAN 2022 tests and all year levels available in a school. The regression analysis shows that 73% of variance in school performance is accounted for by ICSEA.

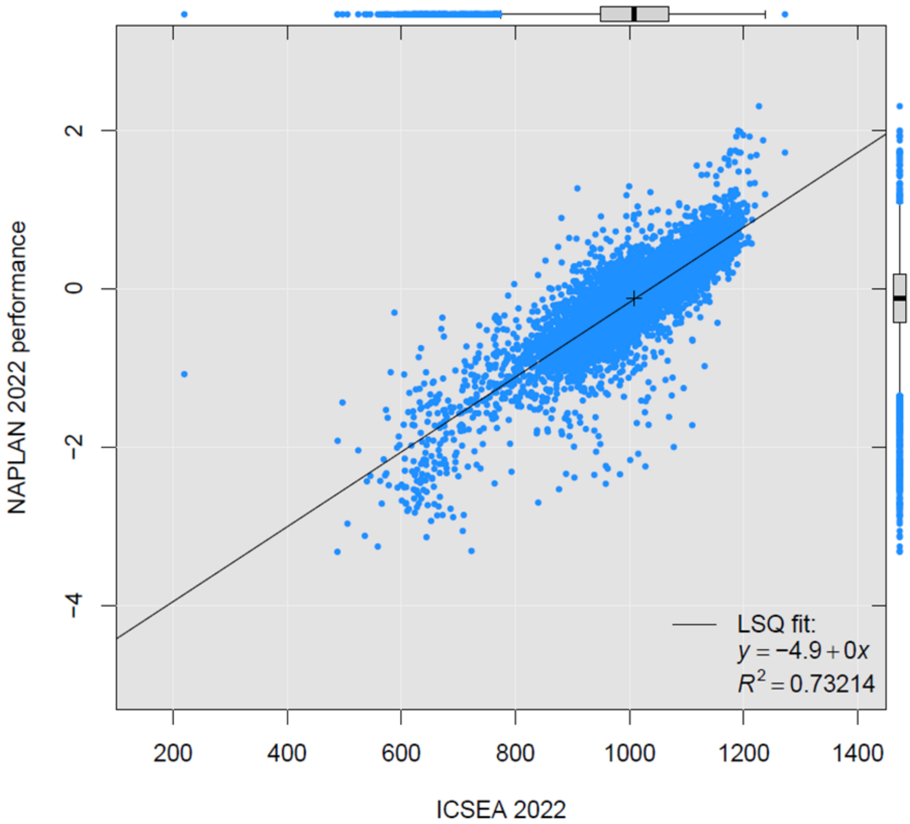

Figure 8: ICSEA 2022 compared with NAPLAN performance

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 2022. The relationship is similar to that exhibited in previous years: positively correlated, but not linearly.

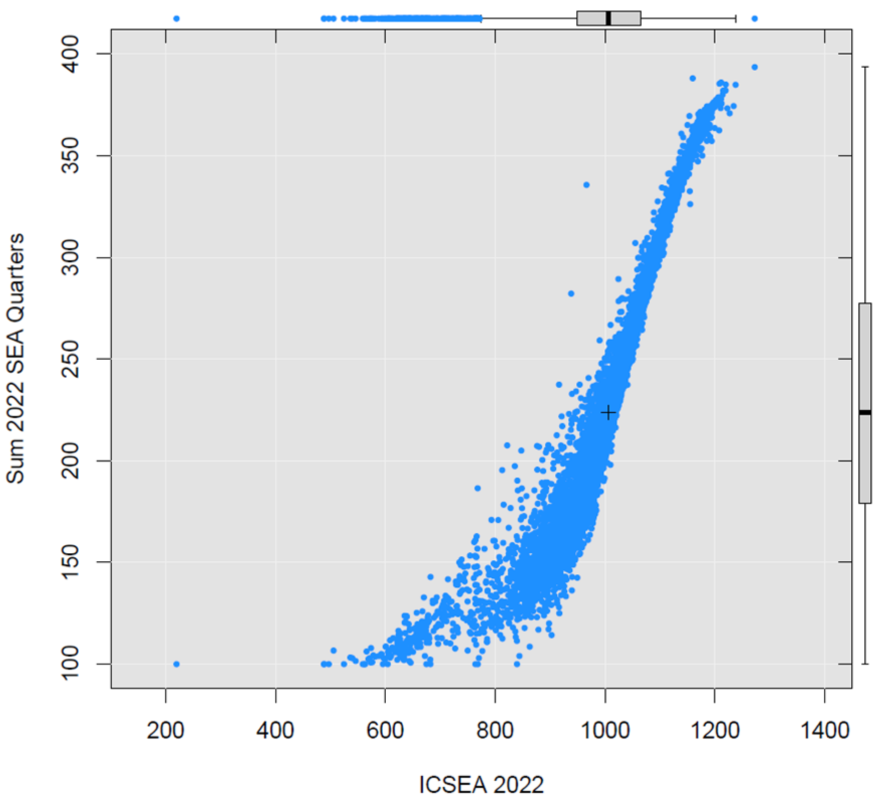

Figure 9: ICSEA 2022 compared with sum of SEA quarters

References

Adams RJ, Wu ML, Cloney D, Berezner A and Wilson M (2020) *ACER ConQuest: Generalised Item Response Modelling Software* (Version 5.29) [Computer software], Australian Council for Educational Research, https://www.acer.org/au/conquest, accessed 16 September 2024.

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.