



Australian Government

Department of Social Services

Australian Institute of Family Studies

Growing Up in Australia:
The Longitudinal Study of Australian Children (LSAC)

LSAC Technical Paper No. 16



The longitudinal study of Australian children

Using *My School* data in the
Longitudinal Study of Australian Children

Kalyca Baker, Brigit Maguire,
Galina Daraganova and Mark Siphthorp

Australian Institute of Family Studies
July 2016

Acknowledgements

This report makes use of data from *Growing Up in Australia: the Longitudinal Study of Australian Children* (LSAC). LSAC is conducted in partnership between the Department of Social Services (DSS), the Australian Institute of Family Studies (AIFS) and the Australian Bureau of Statistics (ABS), with advice provided by a consortium of leading researchers. This report also makes use of *My School* data, which were provided by the Australian Curriculum, Assessment and Reporting Authority (ACARA). Findings and views expressed in this publication are those of the individual authors and may not reflect those of AIFS, DSS, ABS or ACARA.

For more information:

National Centre for Longitudinal Data
Policy Evidence Branch
Australian Government Department of Social Services PO Box 7576
Canberra Business Centre ACT 2610

Email: NCLD@dss.gov.au

Contents

Acknowledgements	ii
List of Tables	2
List of Figures	2
List of Shortened Forms	3
1 Introduction	4
1.1 Growing Up in Australia: The Longitudinal Study of Australian Children	4
1.2 <i>My School</i> data	4
1.3 Importance of linking <i>My School</i> data to LSAC	4
1.4 Structure of this report	5
2 Data linkage and data structure	6
2.1 Linkage and matching process	6
2.2 Data structure	7
3 Confidentialisation	10
4 <i>MySchool</i> data in LSAC	11
4.1 LSAC representativeness of Australian schools	11
4.2 Frequencies for <i>My School</i> variables in LSAC	11
5 Data considerations	25
5.1 Conclusion	27
Appendix A: <i>My School</i> data linkage matching process	28
Appendix B: <i>My School</i> data structure	30
References	35

List of Tables

Table 1:	Total number of schools that have matched data for LSAC children, by Year level and calendar year	12
Table 2:	Number and percent of LSAC children with matched school information, by cohort and by wave	13
Table 3:	School sector, by year for NSW, VIC, QLD, SA and WA	15
Table 4:	School sector, by year for TAS, ACT, NT and Australia (Total)	16
Table 5:	Government and non-government schools, by school type, by year	18
Table 6:	Number of LSAC single-sex and co-educational schools, by school type, by year	19
Table 7:	Average student to teacher ratios in LSAC schools and all schools, by school type for 2008	20
Table 8:	Average student to teacher ratios, by school type for 2009	21
Table 9:	Average student to teacher ratios, by school type for 2010	21
Table 10:	Average student to teacher ratios, by school type for 2011	22
Table 11:	Average LSAC school income, by school type and year	22

List of Figures

Figure 1:	<i>My School</i> data linkage matching process	6
Figure 2:	LSAC distribution of the ICSEA for schools, by year	23

List of Shortened Forms

ABS	Australian Bureau of Statistics
ACARA	Australian Curriculum, Assessment and Reporting Authority
AIFS	Australian Institute of Family Studies
DSS	Department of Social Services
FTE	Full-time equivalent
ICSEA	Index of Community Socio-educational Advantage
LSAC	Longitudinal Study of Australian Children
NAPLAN	National Assessment Program Literacy and Numeracy
SEIFA	Socio-Economic Indexes for Areas

1 Introduction

1.1 Growing Up in Australia: The Longitudinal Study of Australian Children

Growing Up in Australia: the Longitudinal Study of Australian Children (LSAC) is Australia's first nationally representative longitudinal study of child development. The study provides policymakers and researchers with access to quality data about children's development within the current economic, social and cultural environment. The study commenced in 2004 with two cohorts—approximately 5,000 children who were 4 to 5 years old and approximately 5,000 children who were 0 to 1 year old. Children's families were randomly sampled from all Australian states and territories, excluding the most remote areas. Children and their families have been visited every two years since 2004, and data are collected by face-to-face interview, paper questionnaire, computer-assisted self-interview, and direct assessments of the children. For more information about the study, see <http://www.growingupinaustralia.gov.au/>. The study is conducted in partnership between the Department of Social Services (DSS), the Australian Institute of Family Studies (AIFS) and the Australian Bureau of Statistics (ABS).

1.2 *My School* data

Every year since 2008, most Australian students in Years 3, 5, 7 and 9 complete National Assessment Program Literacy and Numeracy (NAPLAN) testing, which assesses students in reading, writing, language conventions and numeracy. *My School* data include NAPLAN information at the school level as well as school characteristics and finances. For example, the NAPLAN means for domains such as reading are provided as an average for the whole **school**, for that Year level and testing year. This is different to NAPLAN information that is provided at the individual **student** level for the Year level and calendar year (see Daraganova, Edwards, & Siphthorp, 2012 for more information about NAPLAN data). School level NAPLAN scores, and other information about the schools participating in NAPLAN, is available for almost 10,000 Australian schools on the *My School* website (www.acara.edu.au). The *My School* data linked to LSAC include detailed information about school performance in NAPLAN and school demographics (e.g., the school type, student population, staff numbers and financial information). Both NAPLAN and *My School* are administered by the Australian Curriculum Assessment and Reporting Authority (ACARA).

1.3 Importance of linking *My School* data to LSAC

Linking key sources of information to LSAC data creates rich contextual information that enables greater detailed analysis to be conducted concerning a large sample of Australian children. Linking relevant aspects of the *My School* data to the LSAC data provides LSAC data users with valuable information about children's educational environments. This is important because children's school environments may have an important influence on development (Hattie, 2009).

Various school characteristics have been investigated in relation to student outcomes. For example, Greenwald, Hedges, and Laine (1996) found that increasing per-student expenditure, particularly when it was directed towards selecting and retaining experienced teaching staff, increased student achievement. Further, when co-educational and single-sex maths classes were compared, the gains in maths achievement were not statistically different between the two class types (Marsh & Rowe, 1996; Rowe, 1988).

My School data provide an opportunity for deeper investigation of school variables in relation to child development than the LSAC data previously permitted. *My School* data have also been linked to the NAPLAN data, enabling the investigation of children's performance at the individual level to be seen in the wider context of the school level. For example, data users could investigate LSAC children's NAPLAN performance relative to that of their peers (Australian children), while controlling for the socioeconomic status of the school they attend.

1.4 Structure of this report

Section 2 of the report describes the linkage process and the structure of the *My School* data linked to the LSAC data. Section 3 explains how the data were confidentialised, and Section 4 describes the distribution of schools in the LSAC sample. Section 5 outlines some considerations about the data that researchers using the *My School* data linked to the LSAC data file should be aware of.

2 Data linkage and data structure

2.1 Linkage and matching process

This section describes the process undergone to link the school level data from *My School* to the LSAC data and the structure of the data file. Three organisations contributed to this process: ACARA, AIFS and ABS. The matching process is depicted in Figure 1. This process is followed to ensure that the linkage procedure protects respondent anonymity and that personal information, such as the name of the school the child attends, is not disclosed. A more detailed written explanation of the matching process is provided in Appendix A.

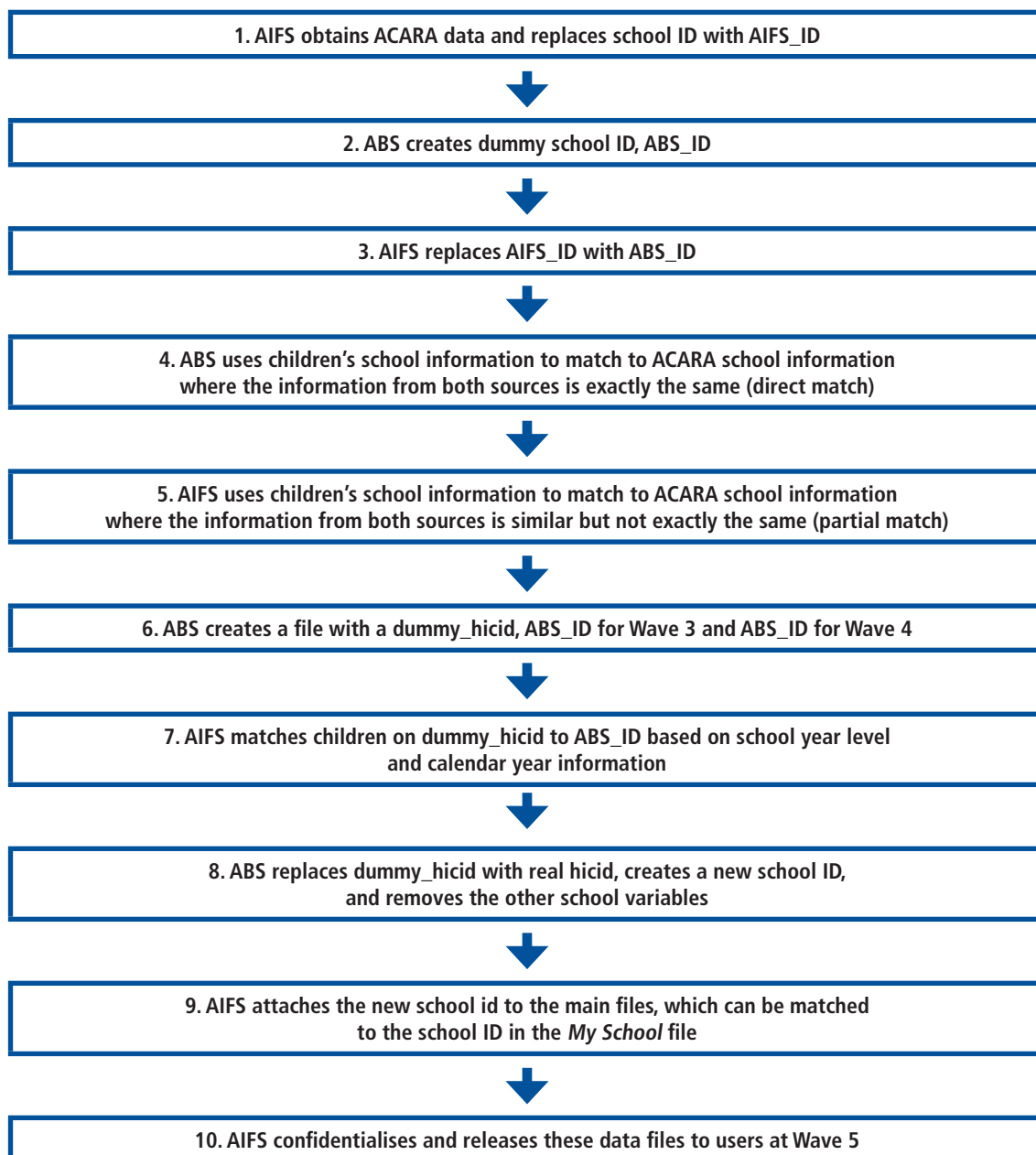


Figure 1: *My School* data linkage matching process

Match type

Direct match

In the first instance, direct matches were made where possible (see step 4 of Figure 1). In these cases, all the details of the school provided in the LSAC interview (LSAC listed schools) corresponded perfectly with the ACARA school information (ACARA listed schools). Matches were based on the school name and street address (suburb, state and postcode, etc.)

Partial match

If a direct match was not possible, a case-by-case investigation was undertaken to see whether the data were similar enough to match. If the differences were minor and it was clear the school was the same, corrections were made to the LSAC listed school (see step 5 of Figure 1) in order for the information to be a complete match to the ACARA listed school. For example, if only the street number was missing, it was copied from the ACARA list into the LSAC list. Another example could be where there was a spelling error in the school name provided in the interview, but the address matched an ACARA listed school, and the school name in the LSAC list was replaced with the correct name from the ACARA list.

In some instances, where the difference was major, a match was still possible and the data could be corrected. In other cases, the data were deemed unmatchable. For example, where the postcode was missing but only one school with the provided name exists in the provided state, the postcode would have been copied into the LSAC list (see step 5 of Figure 1). However, if there was more than one school with a particular name, the data would not have been able to be matched.

No match

Some cases could not be matched through the above procedure. For example, if the school information in the LSAC list did not sufficiently resemble any school on the ACARA list, no match could be made. Similarly, if there was no school information provided at the LSAC interview, no match could be made. If the school information was partial, and the details that were provided were not sufficient to establish one school from the ACARA list, no match was made.

Between wave match/no match

LSAC interview data are collected biennially (e.g., 2008, 2010), so interview information about the child's school is not available for between-waves years (e.g., 2009, 2011). However, longitudinal information allows children to be matched to schools for the between-waves year when there is a reasonable assumption that the child has not changed schools. For example, if the child was matched to the same *school_id* for two consecutive waves, then the same *school_id* was assigned for the between-waves year. Further, if the latest wave interview indicated that the child had not changed schools since the previous wave interview, then the *school_id* from the previous wave was assigned to the between-waves year. Where the child is known to have moved schools between waves, however, no match was made for the between-waves year, because there is uncertainty about which *school_id* should be assigned.

2.2 Data structure

The LSAC *My School* data are stored as a separate data file to the main LSAC data file. Each case represents an individual school at each Year level at each year of NAPLAN testing (when data are available). This means that each school is repeated in the LSAC *My School* data file for each Year level and for each year that NAPLAN data are available for that school (for example, the data file includes a row for an individual school for Year 3 in 2008, 2009, 2010, and so on, and for Year 5 in 2008, 2009, 2010, and so on).

The variable *school_id* is used to match the individual school in the *My School* data to the individual study child in the main LSAC data file. All data from schools that have at least one LSAC study child attending in one of the NAPLAN testing years are included on the LSAC *My School* data file. For example, even if a school has an LSAC child attending in 2008 only, the data file includes data from all Year levels at that school, in 2008 and in all subsequent years (where data are available for the school).

This section describes some of the key variables in detail, and all variables on the data file are summarised in Appendix B and in the Data Dictionary (<http://www.growingupinaustralia.gov.au/data/datadict/index.html>). Note that Appendix B provides information for variables found in the In Confidence data files, and some will be deleted from the General Release files (see below for the confidentialisation procedure). Many of these variables are also explained on the ACARA and *My School* websites (www.acara.edu.au and www.myschool.edu.au), and the descriptions provided here are based on information available on those websites. All variables are reported at the school level, usually for the relevant Year level and test year.

In the instance where children had moved schools between waves, and no match was made, the file structure allows users to make their own decision as to which information they wish to link the child to, depending on the circumstance. For example, the data user can choose to use the school data provided in the previous wave or the following wave for the child's school information for the between-waves year. Having the main data file, NAPLAN data file and *My School* data file as separate files also allows for greater flexibility in analyses. For example, LSAC students can be investigated in comparison with their peers at the individual (NAPLAN) or school (*My School*) level by year, or by calendar year, depending on the research question.

Each of the key variables listed below is provided in the In Confidence file for each of the five NAPLAN assessment domains and for each Year level.

The variable naming is in the format *y#_letter_statistic*, where *y#* represents the Year level (3, 5, 7 or 9), a single letter represents the domain (R=reading, W=writing, S=spelling, G=grammar and punctuation, and N=numeracy), and the statistic represents the type of value presented. For example, *y3_R_mean* is the school's average NAPLAN score for Year 3 reading. For the mean score variables, the single letter can also be prefixed with a combination of letters. These represent Similar Schools (SS), Similar Students (Sim_stud) and All Schools (AS). For example, *y5_SS_W_mean* is the average NAPLAN score for Year 5 Writing for schools that have similar Index of Community Socio-educational Advantage (ICSEA) scores.

Mean scores—The school's mean scores on each testing domain, for the specific Year level in the relevant year of NAPLAN testing.

Assessed percentage—The percentage of students with a reported result. This can differ from the participation rate (which may include exempt, absent and withdrawn students) and the number of eligible children (which is the number of children in the Year level as reported by the school).

Exempt percentage—The percentage of total students who were exempt from completing NAPLAN. This includes students who have significant intellectual or functional disabilities and students from a non-English-speaking background who have been learning English in Australia for less than one year. However, the student's parent can still choose for them to participate.

Absent percentage—The percentage of students who did not complete NAPLAN because they were not present at school on the day of the test or were not able to sit the test (e.g., because of accident or illness).

Withdrawn percentage—The percentage of students whose parents or carers withdrew them from NAPLAN testing (e.g., because of religious or philosophical objections).

Other variables include:

Enrolments—The number of students who attend the school in the specific test year (reported for the total and for males and females separately), including full-time and part-time enrolments.

Full-time equivalent (FTE) enrolments—the sum of the proportion of full-time workload for each student attending the school. Each part-time student is counted as a proportion of the full-time enrolment (e.g., a half-time enrolment is 0.5 FTE).

Attendance rate—The attendance rate for the school as a whole in the specific test year.

Teaching staff—The total number of teaching staff at the school in the specific test year.

Full-time equivalent teaching staff—the total level of staff resources used, where full-time teaching staff are counted as 1.0 and part-time staff are counted as a proportion of full time, e.g., 0.5.

Non-Teaching staff—The total number of non-teaching staff at the school in the specific test year.

Full-time equivalent non-teaching staff—the total level of staff resources used, where full-time non-teaching staff are counted as 1.0 and part-time staff are counted as a proportion of full time, e.g., 0.5.

Total gross recurrent income—the amount of recurrent income received from fees, parent contributions, private sources, and from government (national and state), excluding Government Capital Grants.

Total net recurrent income—the amount of the total gross income that is available for recurrent purposes.

Cohort range—This variable is used to distinguish a subsample of the total sample to allow longitudinal comparisons, such as gain in scores over time. For example, children who participated in Year 3 NAPLAN, then Year 5 NAPLAN two years later, would be in the 3–5 cohort. This group would exclude children who, for example, repeated the Year 3 test or missed one of the tests. Similarly, the 5–7 cohort would be comprised of children who participated in consecutive Year 5 and Year 7 NAPLAN tests, and the 7–9 cohort would be comprised of children who participated in consecutive Year 7 and Year 9 NAPLAN tests.

3 Confidentialisation

The Unconfidentialised (In Confidence) file was released to DSS-approved data users unmodified. The *My School* data are subject to more a vigorous confidentialisation process than other LSAC data because some of the information is publicly available on the *My School* website. Although LSAC data users sign licence agreements stipulating that they cannot attempt to identify respondents, we also have an obligation not to release data in a way that makes identification easier. The Confidentialised (General Release) file was modified before being released to DSS-approved users in order to confidentialise information using the following three methods.

Deletion

Information that was considered potentially sensitive was deleted from the file. For example, some of the more detailed financial information about schools was deleted.

Top and bottom coding

Rather than being deleted completely, some variables are top coded and/or bottom coded for the Confidentialised file. This refers to recoding outlying values to a less extreme value (either from the top of the values, the bottom of values, or both). For example, the variables that contain *similar schools* scores and *similar students* scores are top and bottom coded. The school staff numbers, enrolment numbers and student post-school destinations were also top and bottom coded.

Rounding

The percentage of children who were assessed for each test is rounded to the nearest 5 per cent, so that schools are not identifiable, because the percentage assessed is publicly available on the *My School* website.

For more information about which variables are modified, see Appendix B and the *My School* and NAPLAN tabs of the Data Dictionary (<http://www.growingupinaustralia.gov.au/data/datadict/index.html>).

4 *MySchool* data in LSAC

This section describes the distribution of schools in the LSAC sample by key school characteristics and compares these to national data where available. All LSAC results presented in this section were produced using the In Confidence file. All the variables used here are available in both the In Confidence and General Release files; however, the frequencies presented will differ in the General Release file for the income variables, the ICSEA, and the student to teacher ratios, which have been top and bottom coded and rounded for confidentialisation purposes. For more information about the variables, see Appendix B and the Data Dictionary.

4.1 LSAC representativeness of Australian schools

LSAC studies Australian children over time, and the sample is nationally representative of Australian children of **particular age groups in 2004**. Specific weights have been assigned to each child in the sample and can be used in analysis to produce nationally representative estimates. Analysis using LSAC data should utilise these weights where possible, because they compensate for sampling bias (such as low numbers in rural areas) and to account for retention bias. However, LSAC was designed to be two nationally representative birth cohorts and, given that school starting ages vary from state to state, and parents also have choice in when they send their children to school (see Edwards, Taylor, & Fiorini, 2011), LSAC is not a nationally representative grade cohort of children (also see Daraganova et al., 2013). Therefore, it is unrealistic to expect that the schools that the children in LSAC attend would be nationally representative. *My School* data have been linked to LSAC data in order to enhance the understanding of the LSAC children in the context of their school, rather than to investigate Australian schools per se.

This report provides information at the school level and describes LSAC schools, so that users can see what *My School* data are available. National data are presented in this report to provide some broader context to the LSAC results. The fact that the LSAC school sample was not designed to be nationally representative of Australian schools is important to keep in mind when interpreting comparisons made throughout this report. Any differences between LSAC and national schools are not to be considered a flaw of the LSAC study or sample, although they are noteworthy and may limit researchers' ability to make generalisations about Australian schools.

It is also important to note that weights have not been applied to the analysis in this report. The reason for this is that the analysis throughout the report is performed at the school level (based on *school_id*), whereas weights are available only at the child level (based on *bicid*). More than one child can attend the same school. Therefore, it was more appropriate to conduct analysis at the school, rather than the child, level. Weights are not available to apply to schools in the same way as they are to apply to children.

4.2 Frequencies for *My School* variables in LSAC

Table 1 shows the numbers of schools on the LSAC *My School* data file that have data for each Year level (3, 5, 7 and 9) at each year (2008, 2009, 2010 and 2011). It should be noted that differences in the total numbers of schools between years does not necessarily reflect any change in the total number of schools taking part in NAPLAN testing and data collection—changes in the numbers may also reflect schools opening, closing, combining or splitting between years. It should also be noted that there can be slight differences between the number of schools for each type of assessment (reading, writing, spelling, grammar and punctuation and numeracy) within the same year and Year level. This occurs when NAPLAN data are available for a school on one type of assessment (e.g., numeracy) and not another assessment (e.g., writing). Note that the ACARA data are continually updated and may have changed since AIFS received them. There are fewer secondary results, as fewer LSAC children were in Secondary schools than in Primary schools during Waves 1 to 5.

Table 1 counts each school multiple times when data are available in more than one calendar year and for more than one Year level. There are 3,594 schools for 2008, 3,699 for both 2009 and 2010, and 3,600 for 2011, and most schools have data for more than one calendar year.

Table 1: Total number of schools that have matched data for LSAC children, by Year level and calendar year

Year	Year 3			Year 5			Year 7			Year 9				
	R	W	S & G	R	W	S & G	R	W	S & G	R	W	S & G	N	
2008	3,392	3,389	3,391	3,403	3,405	3,405	3,403	3,403	3,403	1,609	1,609	1,609	1,608	550
2009	3,391	3,392	3,393	3,422	3,421	3,421	3,419	3,419	3,419	1,590	1,590	1,591	1,591	556
2010	3,400	3,399	3,402	3,437	3,436	3,437	3,441	3,441	3,441	1,573	1,573	1,573	1,573	567
2011	3,400	3,397	3,398	3,430	3,428	3,429	3,428	3,428	3,428	1,584	1,583	1,583	1,584	568

Note: R = reading, W = writing, S & G = spelling, grammar and punctuation, N = numeracy.

The following Tables count each school only once for each year, as the information being reported applies to the whole school and is not specific to a particular Year level. The total number of schools in the following Tables may differ from the total number of available schools in the data set, due to missing data on key variables. Tables 3 and 4 show the proportion of Government, Catholic and Independent schools in the LSAC dataset by state for each year. Some states have a higher proportion of Government schools compared with other states. For example, the Northern Territory and Tasmania have a higher proportion of Government schools (around 70 per cent), compared with South Australia and the ACT (around 60 per cent).

Table 2 shows the number of children who do not have school IDs for use throughout this report. As demonstrated, the majority of the B cohort children at Wave 3 do not have school IDs. This is largely explained by the age of these children. At the time of Wave 3 collection, these children are between 4 and 5 years old. They do not have school information available to create a school ID, because they do not yet attend school. However, for the K cohort, who are between 8 and 9 years old at the time of Wave 3 collection, less than 2 per cent of the sample are missing school IDs. By Wave 4, both cohorts only have approximately 2 per cent of cases with missing school IDs. Of those missing school IDs, some do not attend school, some are home schooled, and some did not have sufficient school information to derive a school ID. Given that the number of those missing school IDs is very small ($n = 51-72$ for Waves 4 B cohort and 3 & 4 K cohort), any analysis of demographic characteristics that are associated with non-response would be unreliable.

Table 2: Number and percent of LSAC children with matched school information, by cohort and by wave

	Missing	Wave 3	Wave 4
B Cohort	# Children matched	743	4174
	# Children unmatched	3643	68
	% Children matched	16.94%	98.40%
K Cohort	# Children matched	4239	4072
	# Children unmatched	92	97
	% Children matched	97.88%	97.67%
Both cohorts	# Children matched	4982	8246
	# Children unmatched	3735	165
	% Children matched	57.15%	98.04%

Note: B Cohort children are aged 4 to 5 at Wave 3, hence may not be at school.

Tables 3 and 4 show the national proportions of Government, Catholic and Independent LSAC schools by state from 2009 onwards, as reported in the publicly available ACARA reports, *National Report on Schooling* (ACARA, 2011b, 2012b, 2013b). Ninety-five per cent confidence intervals are presented in brackets for LSAC school proportions only. The number of schools in both government and non-government sectors has remained stable across the four years for all states for both the LSAC and the national school samples.

Catholic schools

The proportion of Catholic schools for South Australia, Western Australia and the Total schools (all state schools combined) was lower for schools nationally than the lower bound of the confidence interval for LSAC across all years. However, the difference in the Catholic school proportions between LSAC and the national sample was minor (between 1 and 2 per cent), and for most states the national sample fell within the LSAC confidence intervals. This indicates that the LSAC Catholic school sample is generally nationally representative.

Government schools

The LSAC school sample proportions of Government schools were significantly lower than the national proportion of all Australian schools across most states (New South Wales, Victoria, Queensland, South Australia and Western Australia) and Total schools (all state schools combined) across all years. Therefore, the LSAC sample under-represents Government schools when compared with the national proportions.

Independent schools

Conversely, the LSAC school sample proportions of Independent schools were significantly higher than the actual proportion of Independent schools across most states (New South Wales, Victoria, Queensland and South Australia) and Total schools (all state schools combined) across all years. Therefore, the LSAC sample over-represents Independent schools when compared with the national proportions.

Table 3: School sector, by year for NSW, VIC, QLD, SA and WA										
	NSW		VIC		QLD		SA		WA	
	Nat	LSAC	Nat	LSAC	Nat	LSAC	Nat	LSAC	Nat	LSAC
2008										
Government	--	66%	--	64%	--	66%	--	62%	--	65%
Catholic	--	20%	--	23%	--	19%	--	18%	--	21%
Independent	--	15%	--	13%	--	14%	--	20%	--	15%
N	--	1,108	--	855	--	737	--	266	--	390
Total	--	100%	--	100%	--	100%	--	100%	--	100%
2009										
Government	70%	66% (63–68)	69%	64% (60–67)	73%	66% (63–70)	75%	62% (56–68)	72%	65% (60–69)
Catholic	19%	20% (17–22)	21%	23% (21–26)	17%	19% (17–22)	13%	18% (14–23)	15%	21% (17–25)
Independent	11%	15% (13–17)	9%	13% (11–15)	10%	14% (12–17)	12%	20% (15–25)	13%	15% (11–18)
N	3,097	1,109	2,279	855	1,710	737	787	268	1,067	390
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
2010										
Government	70%	66% (63–68)	69%	64% (60–67)	73%	66% (63–70)	75%	62% (56–68)	72%	65% (60–69)
Catholic	19%	20% (17–22)	22%	23% (21–26)	17%	19% (17–22)	13%	18% (14–23)	15%	21% (17–25)
Independent	11%	15% (13–17)	10%	13% (11–15)	10%	14% (12–17)	12%	20% (15–25)	13%	15% (11–18)
N	3,092	1,109	2,251	861	1,702	738	775	268	1,065	390
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
2011										
Government	70%	66% (63–68)	69%	63% (60–67)	72%	66% (63–70)	74%	62% (56–68)	72%	65% (60–69)
Catholic	19%	20% (17–22)	22%	23% (21–26)	17%	19% (17–22)	14%	18% (14–23)	15%	21% (17–25)
Independent	11%	15% (13–17)	9%	13% (11–15)	10%	14% (12–17)	12%	20% (15–25)	13%	15% (11–18)
N	3,097	1,109	2,234	860	1,708	738	752	267	1,071	390
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Note: -- indicates national comparisons are not available.

Table 4: School sector, by year for TAS, ACT, NT and Australia (Total)								
	TAS		ACT		NT		Total	
	Nat	LSAC	Nat	LSAC	Nat	LSAC	Nat	LSAC
2008								
Government	--	68%	--	62%	--	71%	71%	65% (63–67)
Catholic	--	17%	--	27%	--	13%	18%	20% (19–22)
Independent	--	15%	--	10%	--	15%	11%	15% (13–16)
N	--	109	--	77	--	52	9,562	3,594
Total	--	100%	--	100%	--	100%	100%	100%
2009								
Government	76%	68% (59–77)	65%	62% (52–73)	81%	71% (59–83)	71%	65% (63–67)
Catholic	14%	17% (10–25)	24%	27% (17–37)	8%	13% (4–23)	18%	20% (19–22)
Independent	11%	15% (8–21)	11%	10% (4–17)	11%	15% (6–25)	11%	15% (13–16)
N	274	109	127	77	188	52	9,529	3,597
Total	100%	100%	100%	100%	100%	100%	100%	100%
2010								
Government	75%	68% (59–77)	65%	62% (52–73)	81%	71% (59–83)	71%	65% (63–67)
Catholic	14%	17% (10–25)	24%	27% (17–37)	8%	13% (4–23)	18%	20% (19–22)
Independent	11%	15% (8–21)	11%	10% (4–17)	11%	15% (6–25)	11%	15% (13–16)
N	268	109	127	77	188	52	9,468	3,604
Total	100%	100%	100%	100%	100%	100%	100%	100%
2011								
Government	75%	67% (58–76)	66%	62% (52–73)	81%	71% (59–83)	71%	65% (63–66)
Catholic	15%	18% (11–25)	23%	27% (17–37)	8%	13% (4–23)	18%	20% (19–22)
Independent	11%	15% (8–22)	11%	10% (6–17)	11%	15% (6–25)	11%	15% (13–16)
N	255	107	128	77	190	52	9,435	3,600
Total	100%	100%	100%	100%	100%	100%	100%	100%

Note: -- indicates national comparisons are not available.

Table 5 shows the number of Government, Catholic and Independent LSAC schools by school type for each year. Combined schools are those that have both Primary and Secondary students in the same school. Most of the LSAC Combined schools are Independent schools (72 per cent). Most of the LSAC Primary schools are Government schools (74 per cent). Currently, there are too few secondary-only schools to report meaningful comparisons for those school types. As the LSAC children move into Secondary schools in the following waves, we will have more information on these schools. There are very few LSAC Special schools.

Table 5 also shows the national proportions of school sector by school type (ACARA, 2011b, 2012b, 2013b). Ninety-five per cent confidence intervals are presented in brackets for LSAC school proportions only. The proportion of LSAC and national schools in both government and non-government sectors has remained stable across the four years for all school types.

Catholic schools

The proportion of Catholic Combined schools and Catholic Secondary schools was similar for the LSAC and national samples across all years, and the LSAC schools can be regarded as representative of schools nationally of this type. However, the proportion of Catholic Primary and Total schools (all school types combined) was slightly lower for national schools than the lower bound of the confidence interval for the LSAC sample, across all years. This indicates that the LSAC Catholic Primary school sample is statistically different from the national sample and is not necessarily representative. There are no LSAC Catholic Special schools, compared with 6 per cent for national schools. However, this is to be expected, given the small number of Special schools in the LSAC sample.

Government schools

The LSAC school sample proportions of Government schools were slightly lower than the national proportion of all Australian schools across all school types and all years. The only exception to this is Government Special schools, where the LSAC proportions were higher than the national. However, there are very few Special schools in the LSAC sample.

Independent schools

The LSAC school sample proportions of Independent schools were significantly higher than the national proportion for Combined schools and Total schools (all school types combined). The proportion of Independent Special schools was lower for the LSAC sample, compared with national proportions. However, again, there are very few of this type of school in LSAC. The Independent schools proportions for Primary and Secondary were similar to national proportions and can be considered generally representative of these school types.

These results reiterate the findings above, that the LSAC sample is not entirely representative of national schools, with a slight under-representation of Government schools in LSAC.

Table 5: Government and non-government schools, by school type, by year										
	Combined		Primary		Secondary		Special		Total	
	Nat	LSAC	Nat	LSAC	Nat	LSAC	Nat	LSAC	Nat	LSAC
2008										
Government	--	19%	--	74%	--	33%	--	95%	71%	65% (63–67)
Catholic	--	9%	--	23%	--	48%	--	0%	18%	20% (19–22)
Independent	--	72%	--	3%	--	19%	--	5%	11%	15% (13–16)
N	--	578	--	2,957	--	21	--	38	9,562	3,594
Total	--	100%	--	100%	--	100%	--	100%	100%	100%
2009										
Government	39%	19% (15–22)	77%	74% (72–75)	72%	35% (14–56)	81%	95% (88–100)	71%	65% (63–67)
Catholic	11%	10% (7–12)	19%	23% (21–24)	22%	45% (23–67)	6%	0%	18%	20% (19–22)
Independent	50%	72% (68–75)	4%	3% (3–4)	6%	20% (2–38)	13%	5% (0–12)	11%	15% (13–16)
N	1,261	581	6,414	2,958	1,439	20	415	38	9,529	3,597
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
2010										
Government	39%	18% (15–22)	77%	74% (72–76)	73%	33% (13–54)	80%	95% (88–100)	71%	65% (63–67)
Catholic	12%	10% (7–12)	19%	23% (21–24)	22%	43% (22–64)	6%	0%	18%	20% (19–22)
Independent	50%	72% (68–76)	4%	3% (3–4)	5%	24% (6–42)	14%	5% (0–12)	11%	15% (13–16)
N	1,286	585	6,357	2,960	1,409	21	416	38	9,468	3,604
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
2011										
Government	39%	18% (15–22)	77%	74% (72–76)	73%	35% (14–56)	78%	95% (88–100)	71%	65% (63–66)
Catholic	11%	9% (7–12)	20%	23% (21–24)	22%	45% (23–67)	6%	0%	18%	20% (19–22)
Independent	50%	72% (68–76)	4%	3% (3–4)	5%	20% (2–38)	15%	5% (0–12)	11%	15% (13–16)
N	1,305	591	6,312	2,952	1,396	20	422	37	9,435	3,600
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Note: -- indicates national comparisons are not available. Only total Australian 2008 schools are provided by ACARA. Therefore, there is no comparison for LSAC schools by school type or sector for 2008. The Secondary and Special school results are to be interpreted with caution, as there are so few of these school types in the LSAC sample.

The Special schools are excluded from the frequencies presented in the remainder of this section to ensure confidentialisation, because the numbers are low and because these schools are not necessarily comparable to other school types.

Table 6 shows the number of single-sex and co-educational LSAC schools. There are very few single-sex schools in the LSAC school sample. Of the few schools in the sample that are single sex, most are classified as Combined schools, rather than Primary or Secondary schools.

Table 6: Number of LSAC single-sex and co-educational schools, by school type, by year				
	Combined	Primary	Secondary	Total
2008				
Boys only	41	3	2	46
Girls only	39	0	4	43
Co-educational	414	2,942	9	3,403
Total	494	2,945	15	3,492
2009				
Boys only	41	3	2	46
Girls only	41	0	3	44
Co-educational	413	2,948	10	3,409
Total	495	2,951	15	3,499
2010				
Boys only	42	3	2	47
Girls only	41	0	3	44
Co-educational	419	2,954	11	3,422
Total	502	2,957	16	3,513
2011				
Boys only	42	0	3	45
Girls only	42	3	2	47
Co-educational	423	2,946	10	3,416
Total	507	2,949	15	3,508

Note: ACARA does not report co-educational, single-sex school status, so national comparisons are not available for these variables.

Tables 7 to 10 show the average student to teacher ratios in LSAC and national schools by school type for each year. Student to teacher ratios were calculated by dividing the number of FTE teaching staff by the number of FTE student enrolments. This is not the same as class size, although the values may be closely related. Ratios do not account for class characteristics, such as student age, Year level or subject type. They do not account for specialist or administrative duties.

Data presented in Tables 7 to 10 compare LSAC to national mean ratios (ACARA, 2011b, 2012b, 2013b) by school type and by school sector. Ninety-five per cent confidence intervals are presented in brackets, for the LSAC sample only. For both the LSAC sample and the national sample, ratios remained consistent across all years.

For the Catholic and Independent Primary schools, the national mean ratio was the same as the LSAC mean, or fell within the LSAC confidence interval ratios, across all years. Therefore, these schools are considered to be statistically similar, and the LSAC sample is considered to be nationally representative with respect to teacher to student ratios for these school types.

For the Government Primary schools, however, mean LSAC ratios were somewhat higher than the national average across all years and are considered statistically different to the national schools. These differences were, however, reasonably small, in the order of one ratio point. For example, for 2008 data, the difference between the LSAC and national ratios equates to one extra teacher per 320 students.

For Secondary schools, the mean Government ratios and Independent school ratios were not statistically different for LSAC and national schools across all years. For Catholic Secondary schools, mean LSAC ratios were significantly higher than the national average across all years.

'All schools' ratios include Special schools and Combined schools. This may explain why the LSAC ratio means differ from the national when all schools are totalled. This highlights the need to select specific school types or sectors when conducting analyses with LSAC data, depending on the research topic of interest.

Table 7: Average student to teacher ratios in LSAC schools and all schools, by school type for 2008								
	Government		Catholic		Independent		All schools	
	Nat	LSAC	Nat	LSAC	Nat	LSAC	Nat	LSAC
Primary								
Mean FTE teaching staff	--	23	--	19	--	14	--	22
Mean FTE total staff	--	30	--	25	--	20	--	28
Mean FTE student enrolments	--	375	--	339	--	204	--	361
Mean ratio of students to teaching staff	15.6	16.4 (16.3–16.4)	17.6	17.7 (17.6–17.8)	14.7	14.9 (14.7–15.2)	--	16.6 (16.5–16.7)
Secondary								
Mean FTE teaching staff	--	20	--	58	--	60	--	58
Mean FTE total staff	--	41	--	80	--	77	--	78
Mean FTE student enrolments	--	789	--	576	--	732	--	733
Mean ratio of students to teaching staff	12.3	12.2 (11.6–12.8)	12.8	13.9 (13.6–14.3)	10.5	9.8 (8.6–11)	--	12.5 (12.0–12.9)
All schools	--	16.0 (15.9–16.1)	--	17.4 (17.3–17.5)	--	13.4 (13.2–13.6)	13.9	15.9 (15.8–15.9)

Note: -- indicates national comparisons are not available.

Ratio = Total FTE student enrolments divided by Total FTE teaching staff. Student to teacher ratios are calculated for each school for the total sample of schools, whereas the results presented in Tables 7 to 10 are mean results of teaching, student and ratios, presented as a subsample of the total sample of schools (e.g., for Government schools). Therefore, the mean ratios in the Tables will not equal mean FTE student enrolments divided by mean FTE teaching staff in the Tables.

Table 8: Average student to teacher ratios, by school type for 2009								
	Government		Catholic		Independent		All schools	
	Nat	LSAC	Nat	LSAC	Nat	LSAC	Nat	LSAC
Primary								
Mean FTE teaching staff	--	23	--	19	--	14	--	22
Mean FTE total staff	--	30	--	25	--	20	--	29
Mean FTE student enrolments	--	378	--	339	--	210	--	363
Mean ratio of students to teaching staff	15.5	16.4 (16.3–16.4)	17.6	17.7 (17.6–17.8)	14.8	14.9 (14.7–15.2)	15.8	16.6 (16.5–16.7)
Secondary								
Mean FTE teaching staff	--	58	--	58	--	60	--	58
Mean FTE total staff	--	78	--	81	--	77	--	77
Mean FTE student enrolments	--	801	--	583	--	743	--	733
Mean ratio of students to teaching staff	12.3	12.2 (11.6–12.8)	12.8	13.9 (13.6–14.3)	10.6	9.8 (8.6–11)	12.0	12.5 (12.0–12.9)
All schools	14.0	16.0 (15.9–16.1)	15.1	17.4 (17.3–17.5)	12.2	13.4 (13.2–13.6)	13.9	15.9 (15.8–15.9)

Note: -- indicates national comparisons are not available

Table 9: Average student to teacher ratios, by school type for 2010								
	Government		Catholic		Independent		All schools	
	Nat	LSAC	Nat	LSAC	Nat	LSAC	Nat	LSAC
Primary								
Mean FTE teaching staff	--	23	--	19	--	14	--	22
Mean FTE total staff	--	30	--	25	--	20	--	29
Mean FTE student enrolments	--	382	--	340	--	211	--	367
Mean ratio of students to teaching staff	15.4	16.4 (16.3–16.4)	17.6	17.7 (17.6–17.8)	14.9	14.9 (14.7–15.2)	15.7	16.6 (16.5–16.7)
Secondary								
Mean FTE teaching staff	--	59	--	60	--	59	--	60
Mean FTE total staff	--	79	--	87	--	75	--	81
Mean FTE student enrolments	--	842	--	631	--	723	--	759
Mean ratio of students to teaching staff	12.3	12.2 (11.6–12.8)	12.8	13.9 (13.6–14.3)	10.5	9.8 (8.6–11)	12.0	12.5 (12.0–12.9)
All schools	14.0	16.0 (15.9–16.1)	15.1	17.4 (17.3–17.5)	12.2	13.4 (13.2–13.6)	13.9	15.9 (15.8–15.9)

Note: -- indicates national comparisons are not available.

Table 10: Average student to teacher ratios, by school type for 2011								
	Government		Catholic		Independent		All schools	
	Nat	LSAC	Nat	LSAC	Nat	LSAC	Nat	LSAC
Primary								
Mean FTE teaching staff	--	19	--	14	--	60	--	22
Mean FTE total staff	--	25	--	20	--	77	--	29
Mean FTE student enrolments	--	342	--	210	--	743	--	370
Mean ratio of students to teaching staff	15.3	16.4 (16.3–16.4)	17.5	17.7 (17.6–17.8)	14.8	14.9 (14.7–15.2)	15.6	16.6 (16.5–16.7)
Secondary								
Mean FTE teaching staff	--	59	--	62	--	59	--	60
Mean FTE total staff	--	75	--	83	--	83	--	80
Mean FTE student enrolments	--	723	--	861	--	549	--	750
Mean ratio of students to teaching staff	12.2	12.2 (11.6–12.8)	12.8	13.9 (13.6–14.3)	10.4	9.8 (8.6–11)	12.0	12.5 (12.0–12.9)
All schools	13.9	16.0 (15.9–16.1)	15.0	17.4 (17.3–17.5)	12.1	13.4 (13.2–13.6)	13.8	15.9 (15.8–15.9)

Note: -- indicates national comparisons are not available.

Table 11 shows the average reported LSAC school incomes by sector for 2009 and 2010. Independent schools have the highest average income, with a substantial difference between gross and net incomes. Government and Catholic schools have similar incomes, with a greater difference in gross and net income for Catholic schools, compared with Government schools. National comparisons are not provided for the variables in Table 11 because national data are presented as recurrent income only (gross/net is not specified) and are reported by financial year rather than by calendar year.

Table 11: Average LSAC school income, by school type and year			
	Government	Catholic	Independent
	\$	\$	\$
2009			
Total gross recurrent Income	3,723,168	3,635,881	10,859,055
Total net recurrent income	3,700,545	3,322,023	9,699,425
2010			
Total gross recurrent Income	3,843,545	3,892,349	11,725,905
Total net recurrent income	3,835,750	3,591,415	10,373,715

Note: Financial data are not available for 2008 or 2011. National comparisons are not available for these variables.

ICSEA is a measure of social, community and educational factors that has been calculated by ACARA for each school for each year since 2009. It has a mean of 1000 and a Standard Deviation of 100. Scores range from 500 to 1300, with higher scores representing greater levels of advantage. Each school's score is the average score of all students attending that school for that year (ACARA, 2011a). It allows school information to be viewed within the broader context of the characteristics of children who attend the school, and allows comparisons of schools that have similar ICSEA scores.

The distribution of ICSEA for all Australian schools for 2011 and 2012 (ACARA, 2013c), and Figure 2 displays the distribution of ICSEA for LSAC schools for 2009, 2010 and 2011. 2008 was the first year of collection of ICSEA information. The process was still under development, and the data were incomplete. Data from 2009 were used to impute 2008 values. In the LSAC *My School* data, the 2008 ICSEA values that are provided are exactly the same as the 2009 values. Therefore, 2008 values are not included in Figure 2.

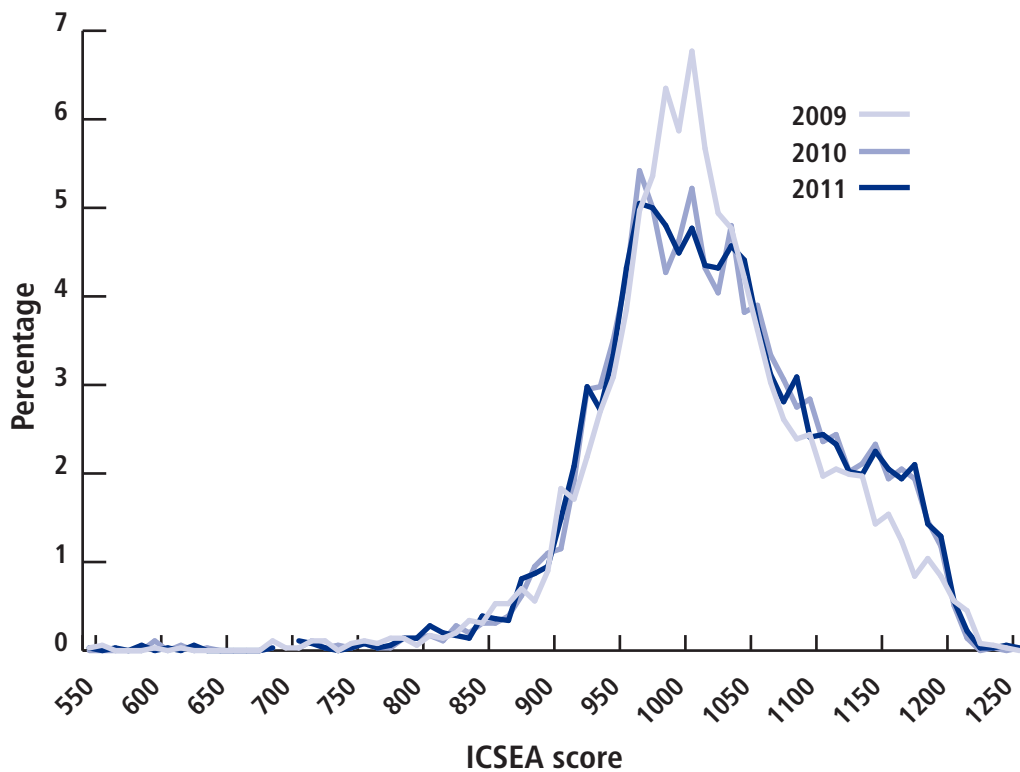


Figure 2: LSAC distribution of the ICSEA for schools, by year

Source: ACARA, 2013c

The distribution of ICSEA for LSAC sample is a similar shape to the national ICSEA distribution for 2011 (ACARA, 2013c, p.10). As with the national ICSEA distribution, the LSAC ICSEA distribution is negatively skewed, with longer tail in the lower advantaged schools, and peak at ICSEA scores of approximately 1000. However, the LSAC school sample appears to have a greater proportion of more advantaged schools in comparison to the national distribution for 2010 to 2011. It can be assumed that the LSAC sample is approximately representative of national schools with regard to ICSEA in 2009 and slightly over-represents more advantaged schools in 2010 and 2011. The LSAC sample has previously been found to be slightly skewed to more educated parents (Daraganova & Siphthorp, 2011), so the higher ICSEA scores are not surprising.

The difference in the distributions from 2009 to 2010 could reflect a change in schools, with LSAC children moving from a lower ICSEA Primary school to a higher ICSEA Secondary school. However, the change could also be due to the retention rate falling over the years for the LSAC study. LSAC retention has been found to be non-random, with a higher proportion of lower socioeconomic families withdrawing from the study in comparison with higher socioeconomic families (Daraganova & Siphthorp, 2011). For those schools that are the same in both years, the composition of students within that school will have changed, and the school's ICSEA will change as a result. Finally, the change could be due to the method of ICSEA calculation. The method by which the ICSEA value is calculated is subject to ongoing review and has changed between collection years (ACARA, 2010, 2011a, 2012a, 2013a).

For example, in 2009, ICSEA was calculated as socio-educational advantage, plus school's remoteness index, plus the percentage of Indigenous enrolments (ACARA, 2011a). In 2009, socio-educational advantage was based on proportional information of parental income, education, occupation, family composition and proportion information of community Indigenous status and internet connection. This information was sourced from the ABS 2006 Census data (ACARA, 2010).

In 2010, ICSEA was calculated as socio-educational advantage, plus school's remoteness index, plus the percentage of Indigenous enrolments, plus the percentage of parents with a combination of language background other than English and education of Year 9 equivalent or below (ACARA, 2011a). In 2010, socio-educational advantage was based on 'direct' and 'indirect' methods. The indirect method is that described for 2009, sourced from the ABS 2006 Census data; however, fewer variables were included. The direct method utilises parental occupation and education information sourced from school records, which were collected directly from parents upon student enrolment. The variables measuring occupation and education in the direct method differ from the ones used in the indirect method. For the majority of cases, socio-educational advantage information was sourced from the direct method (ACARA, 2010).

The calculation equation for ICSEA in 2011 is the same as for 2010 (ACARA, 2012a). There have been subsequent changes to the calculation of later version of ICSEA that are not used in this report. These changes are documented in ACARA (2013c). The changes in ICSEA between 2011 and 2012 to 2013 do not appear to be as substantial as the earlier changes described above. Each year, the proportion of data acquired through the direct method increases, which should improve the data accuracy. The main difference in the 2012 and 2013 calculations is that the language background other than English and education of Year 9 equivalent or below are no longer included. See the considerations section for a more detailed discussion on the issues associated with using the ICSEA variable longitudinally.

5 Data considerations

This section describes some specific constraints of the *My School* data. Users should take these issues into consideration in their analyses and, where appropriate, acknowledge them in their results.

No campus specific data

In the LSAC-released *My School* data, schools that have more than one campus are required to report data at the whole school level, rather than at an individual campus level. Data users will not be able to separate out the *My School* information to the campus level. Since campuses usually have separate geographical locations, and because schools with multiple campuses may differ from single campus schools in other ways, it is important for data users to consider the implications of the aggregated data. This is particularly important where location is relevant to the research question. For example, a multi-campus school's ICSEA score will be based on information from all its campuses, but the suburbs in which the campuses are located could be very different.

This issue also relates to other geographical variables. In the LSAC main data files, there is a Socio-Economic Indexes for Areas (SEIFA) score which ranks the location where the child resides on a level of advantage. Researchers may want to investigate the relationship between the residential advantage score (SEIFA) and the school advantage score (ICSEA). However, because the ICSEA score is generated at the whole school level for multi-campus schools, it is possible that a child could attend a school campus in the suburb where they reside, and still have vastly different SEIFA and ICSEA scores. They may also not reside in the same suburb in which they go to school.

Collection period

The time of year and collection period for LSAC interviews is likely to differ from the period when *My School* data are collected. Data users should take note of when data were collected. For example, the characteristics of the school that the child currently attends, and attended during NAPLAN, cannot be attributed to the classroom behaviour results from previous months, when the child attended a different school. The temporal sequence of when data were collected matters for analysis.

Data users should be aware of the delay in the provision of financial data. Financial data is reported by calendar year to align with the school operating year (1 January to 31 December). All financial data is generally submitted to ACARA by the end of August each year, and then quality assured for publication on the *My School* website in the new calendar year (around late February to early March). Therefore, there is a one-year lag in financial reporting compared to the other data.¹

Limited Secondary schools

At the time of release of the first *My School* data (2014), there were far fewer LSAC children participating in Year 7 or 9 NAPLAN tests. This is because of their age and Year level. This means that there is limited Secondary school representation in comparison with Primary schools. The rates of Primary and Combined schools are not likely to change substantially in the coming waves, because schools are not removed from the data set, even if the child no longer attends the school. Conversely, the number of Secondary schools in the data set will continue to grow as LSAC children transition into Secondary schools. We suggest selecting a subsample of the data, using only the school type of interest, or attempting to control for school type in analyses in an appropriate way.

¹ Government and non-government financial reporting systems do not align to the same reporting period. While non-government school financial reporting to the Commonwealth is based on a calendar year, state and territory government Finance Statistics and State Budgets are based on a financial year. There was a Ministerial decision to report the financial data by calendar year to align with the general school-operating year. All financial data submitted must relate to the 12 calendar months of the year being reported (1 January to 31 December); that is, a normal school year.

Non-matched LSAC children

Not all of the LSAC sample could be matched to *My School* data. Currently, those non-matched cases are very few and are primarily a result of the children not being of school age. The number of cases who have missing school IDs due to other reasons (insufficient school information) is likely to remain stable across waves. These children, as a group, are unlikely to differ substantially from the rest of the sample. However, there were too few cases to support a reliable investigation of whether there were statistically significant differences on various characteristics.

Similarly, users should be alert to the matching method before performing analyses. As discussed above, in some instances, school information was copied from the ACARA file to accomplish a match between LSAC and *My School* data. Even though this was carried out with caution and with the assumption that the information was correct, there remains nonetheless the possibility that the inserted school information was not a true reflection of the school that the child attends.

Non-matching between waves

As stated earlier, if the child is known to have moved schools between waves, no matches between LSAC children and *My School* data were applied to the between-waves year. If the data user wishes to link a school to a child between years, they will have to decide on a methodology that is most appropriate for their research. For example, the user could assign the *school_id* from the previous wave, or assign the *school_id* for the following wave. It is important to take other variables into consideration when making these decisions. For example, when deciding which school to assign if the *school_id* has changed during the Primary school to Secondary school transition period (Years 5 to 8), it will be important to see whether the child was in Year 6, 7, or 8 at the time of the LSAC interview. The state must also be considered, as Queensland, South Australian and Tasmanian students transition to Secondary school later than students in the other states. The data user may wish to apply these linking decisions to all available cases in the same way or to match children to schools on a case-by-case basis. Whatever the approach, researchers should explain their rationale to readers when publishing data.

School change

The *My School* data are longitudinal and therefore are subject to change over time. Some schools that are assessed might undergo substantial changes in the values of key variables used in analyses. For example, campuses may merge and schools may merge. In this instance, ACARA may assign a new school ID or may keep a previous school ID, depending on the individual circumstance. Data users are encouraged to be mindful of this when using the data longitudinally.

Test change

The NAPLAN tests themselves are also subject to change. For example, in 2011, the writing task changed from a narrative writing test to a persuasive writing test. The NAPLAN website suggests that, as performance on the two types of writing test can differ, results should not be compared (<http://www.nap.edu.au/information/faqs/naplan--writing-test.html>). This means that writing scores from 2008 to 2010 should not be compared with results from 2011 (onwards).

Comparing individual schools

While there is some clustering of students attending schools, there are too few to robustly compare individual schools, even though the majority of the sample share schools. For Wave 3 (both cohorts), 72 per cent of the LSAC schools had more than one LSAC child attending. For Wave 4 (both cohorts), 80 per cent of the schools had more than one LSAC child attending. We expect that more children will share schools as children progress to Secondary school, as there are generally fewer Secondary schools to choose from than Primary schools.

ICSEA change

The way ICSEA is calculated is subject to ongoing change (ACARA, 2013c). The different calculation methods reportedly produce highly correlated ICSEA values (ACARA, 2011a, 2012a, 2013a). However, there are nonetheless potential impacts on analysis and subsequent conclusions drawn from these results when using ICSEA longitudinally. It is not possible to discern whether changes in ICSEA over time are due to school composition or the ICSEA calculation method. The difficulty in tracking and quantifying the potential impact is limited because of the magnitude of change and the lack of specific detail about these changes. For example, not only has the equation used to derive ICSEA changed over time, but the variables used in these calculations, as well as the source of these variables, have changed over time. Although LSAC has similar variables to those purportedly used by ACARA, we would require access to the actual data from the component variables that make up the ICSEA scale to be able to investigate the changes more closely.

ACARA advise that ICSEA be used to provide a context for school comparisons² and not for longitudinal analyses. The largest methodological changes in shifting from indirect to direct collection of parental information between 2008–09 to 2010 (see the results section for more information) mean that ICSEA should not be used longitudinally between 2008–09 and 2010. The indirect model used in 2008–09 under-represents the level of disadvantage in Government schools and, at the same time, under-represents the level of advantage in non-government schools when compared with the newer direct method (ACARA, 2010, 2011a). Therefore, we suggest that the ICSEA variable for 2008 and 2009 should be treated as a different measure to the ICSEA variable for 2010.

As the changes beyond 2010 are more minimal, the comparability for ICSEA from 2010 to 2013 is greater. However, data users should note that bias nevertheless exists in the direct method. In particular, the direct method results in more variability over time in the ICSEA for small schools. As there are fewer parents in small schools from whom data can be collected, slight changes in parental data will have a more pronounced effect on the average education and occupation level for the whole school. In contrast, a change in a few parental education and occupation levels is unlikely to change the average value for a large school. ICSEA should therefore be more accurate for large schools than for small schools.

National representativeness

As described in the results section, the LSAC school data are not representative of Australian schools in general. It is not the intent of the LSAC project to have a nationally representative school sample; nor should this be considered a flaw of the study. Researchers should not generalise LSAC school results to the general population of Australian schools.

5.1 Conclusion

In summary, this technical paper has described *My School* data within the context of LSAC data. This included an explanation of the linking process and the structure of the data file. Further, key characteristics of the LSAC schools have been described and compared with national data. These comparisons revealed that the LSAC school sample might not be entirely representative of Australian schools in general. Finally, some important data issues were raised for the consideration of data users. In particular, some implications for analyses were discussed regarding the limited number of LSAC Secondary schools in the sample to date and the longitudinal changes to the ACARA schools, NAPLAN testing conditions and the ICSEA variable.

² 'ICSEA was developed to enable fair and meaningful comparisons of the performance in literacy and numeracy of a given school with that of schools serving students with statistically similar backgrounds. The ICSEA is recalculated on a yearly basis and is not designed for time-series analysis. This data is not recommended to be used for any other purposes' (V. Dao, ACARA, personal communication, November 10, 2014).

Appendix A: *My School* data linkage matching process

Step 1: Obtain ACARA *My School* data

AIFS requested *My School* data through ACARA's formal data request procedure. A Memorandum of Understanding was created and signed by both parties, AIFS and ACARA. ACARA then provided AIFS with a full list of schools and associated data. The data included school location information, school demographics, school NAPLAN data and school financial data.

Step 2: De-identify the *My School* data

It is important that data users are not provided with information that will enable identification of a school and thereby potentially identify an LSAC study child. For this reason, various dummy identifiers replace original identifiers, and other variables were removed from the file. For example, the *ACARA_ID* is available on the *My School* website, so it could not be provided on the released data files. AIFS added a dummy ID variable to the *My School* data and removed the identifiable information about schools from the data file, e.g., School Name, Address, *ACARA_ID*. The identifying variables, along with the *AIFS_ID* variable, were sent to ABS (ACARA master school list).

Step 3: ABS identifier

ABS added a new dummy_id (*ABS_ID*) to the file sent by AIFS. The concordance file (i.e., a file that contains the concordance between the two IDs) is stored by ABS. ABS sent the *ABS_ID* to AIFS, and AIFS then added the *ABS_ID* to the current *My School* dataset and removed the *AIFS_ID*. The *AIFS_ID* was then deleted by AIFS.

Step 4: Identifying *My School* schools attended by LSAC children

ABS used the respondent's identifying variables to match against the ACARA master school list. This data was matched for all children in the LSAC study (see Section 2.1 for more information on matching). After the matching process was completed, the *ABS_ID* of the school was added to the children's records. Where a direct match was not obtained, a file containing a *dummy_school_id* and *dummy_bicid* was sent to AIFS, and they manually reviewed the information collected in the interview to identify any errors which may have prevented a match. Once this process was complete, AIFS sent this file back to ABS, and ABS added the *ABS_ID* to the child's record.

Step 5: Respondent IDs and *ABS_ID*s sent to AIFS

ABS sent a file containing the *bicid*, Wave 3 *ABS_ID*, Wave 4 *ABS_ID*, along with all other relevant variables to AIFS for all children present in any of the waves.

Step 6: Link LSAC and *My School* together

The *My School* file exists as a long file with a separate record for each combination of school, Year, and NAPLAN assessment year. Each school has multiple years, (2008, 2009, etc.) and each year has multiple NAPLAN assessments, 3, 5, 7 and 9. To link the files together, we needed to determine which school the child attended (*ABS_ID* provided by ABS), which Year they were in at school (obtained from the interview) and in what calendar year they were attending that Year (obtained from the interview). Once AIFS established the school ID (*ABS_ID*), Year and calendar year, they were able to match the corresponding row in the *My School* file.

Step 7: Create final school ID variable for release

ABS then assigned a new dummy ACARA ID, named *school_id*, to each school, to be used as the school identifier available for data users. The value of *school_id* will be constant for each school across waves, enabling it to be used longitudinally. New IDs will be created for each wave only for new schools (for example, schools that participate in NAPLAN for the first time in that year). Once each school had been assigned a *school_id*, the school names and location variables were stripped off the file, leaving only the dummy study child ID and the final school identifier.

Step 8: Reassign the study child ID

ABS then replaced the dummy child ID with the real study child ID and provided AIFS with a file now containing only the real study *child_id* and the *school_id*.

Step 9: Confidentialise *My School* data

AIFS then confidentialised the data (see Step 10 for the confidentialisation procedure).

Step 10: Release data

AIFS then added the *school_id* variable to the main data file and created a separate *My School* data file that has only the *My School* information. These files are released to data users.

Using this linkage methodology, ABS is never able to view the *My School* data, and AIFS is never able to access the real child ID in conjunction with the school information, so neither agency can know which child belongs to which *My School* record. The *school_id* is assigned so that data users can apply their own rules to overcome the issue of biannual interviewing and annual attendance at school.

Appendix B: *My School* data structure

Variable Name	Variable Label	Values	Confidentialisation
school_id	School ID	Number	None
School_Post_Code	Post Code	Number	Deleted
Total_Enrolment	Total Enrolments	Number	Top/Bottom Code & Round
FTE_Total_Enrolment	FTE Total Enrolment	Number	Top/Bottom Code & Round
Total_Female_Enrolments	Total Female Enrolments	Number	Top/Bottom Code & Round
Total_Male_Enrolments	Total Male Enrolments	Number	Top/Bottom Code & Round
Teaching_Staff_numbers	Teaching Staff numbers	Number	Top/Bottom Code & Round
FTE_teaching_staff_numbers	FTE teaching staff numbers	Number	Top/Bottom Code & Round
Non_teaching_staff_numbers	Non-teaching staff numbers	Number	Top/Bottom Code & Round
FTE_non_teach_staff_number	FTE non-teaching staff numbers	Number	Top/Bottom Code & Round
School_ICSEA	School_ICSEA	Number	Top/Bottom Code & Round
Number_of_Vet_Enrolments	Number of Vet Enrolments	Number	Top/Bottom Code & Round
Numbers_Vet_Qualifications	Numbers Vet Qualifications	Number	Top/Bottom Code & Round
SBAT	Student Based Apprenticeships and Traineeships	Number	Top/Bottom Code & Round
Senior_Secondary_Cert	Senior Secondary Certificate	Number	Top/Bottom Code & Round
Completed_Senior_Secondary	Completed Senior Secondary	Number	Top/Bottom Code & Round
RI_AG_RecurrentFund	Australian Government recurrent funding_\$ Total	Number	Top/Bottom Code & Round
RI_SG_RecurrentFund	State/Territory Government recurrent funding_\$ Total	Number	Top/Bottom Code & Round
RI_Other_private_sources	Other private sources_\$ Total	Number	Top/Bottom Code & Round
RI_Fees_parental_contrib	Fees, charges and parent contributions_\$ Total	Number	Top/Bottom Code & Round
CE_Other_private_sources	Other_\$ Total	Number	Delete
CE_AG_CapFund	Australian Government capital expenditure_\$ Total	Number	Top/Bottom Code & Round
CE_SG_CapFund	State/Territory Government capital expenditure_\$ Total	Number	Top/Bottom Code & Round
CE_NewSchoolLoans	New school loans_\$ Total	Number	Delete
CE_Feelnc_Alloc_CurrCapProj	Income allocated to current capital projects_\$ Total	Number	Delete
D_Feelnc_Alloc_CurrCapProj	Income allocated to current capital projects_\$ Total	Number	Delete
D_Feelnc_Alloc_FutCapProj	Income allocated to future capital projects and diocesan capital funds_\$ Total	Number	Delete
D_Feelnc_Alloc_DebtServicing	Income allocated to debt servicing	Number	Delete

Variable Name	Variable Label	Values	Confidentialisation
FTE_Funded_Enrolments	Full-time equivalent enrolments relating to recurrent income and capital expenditure	Number	Delete
RI_AG_RecurrentFundPS	Australian Government recurrent funding_ \$ Per student	Number	Delete
RI_SG_RecurrentFundPS	State/Territory Government recurrent funding_ \$ Per student	Number	Delete
RI_Fees_parental_contribPS	Fees, charges and parent contributions_ \$ Per student	Number	Delete
RI_Other_private_sourcesPS	Other private sources_ \$ Per student	Number	Delete
CE_Other_private_sourcesPS	Other_ \$ Total Per Student	Number	Delete
CE_AG_CapFund_PS	Australian Government capital expenditure_ \$ Total Per Student	Number	Delete
CE_SG_CapFund_PS	State/Territory Government capital expenditure_ \$ Total Per Student	Number	Delete
CE_NewSchoolLoans_PS	New school loans_ \$ Total Per Student	Number	Delete
CE_Feelnc_Alloc_CCP_PS	Income allocated to current capital projects_ \$ Total Per Student	Number	Delete
D_Feelnc_Alloc_CurrCapPro_PS	Income allocated to current capital projects_ \$ Per student	Number	Delete
D_Feelnc_Alloc_FutCapPro_PS	Income allocated to future capital projects and diocesan capital funds_ \$ Per student	Number	Delete
D_Feelnc_Alloc_DebtServ_PS	Income allocated to debt servicing (including principal repayments and interest on loans)_ \$ Per student	Number	Delete
RI_TotRecurrentInc	Total gross income (excluding income from government capital grants)_ \$ Total	Number	Top/Bottom Code & Round
RI_TotRecurrentInc_Gov	Total recurrent funding—Government	Number	Delete
RI_TotRecurrentInc_Pri	Total recurrent funding—Private	Number	Delete
CE_TotCapExpend	Total capital expenditure_ \$ Total	Number	Top/Bottom Code & Round
D_TotDeductions	Subtotal_ \$ Total—Deductions	Number	Delete
D_TotNetRecurrentInc	Total net recurrent income_ \$ Total	Number	Top/Bottom Code & Round
D_TotNetRecurrentInc_Private	Total net recurrent income_ \$ Total—Private	Number	Delete
RI_TotRecurrentIncPS	Total gross income (excluding income from government capital grants)_ \$ Per student	Number	Delete
TotRecurrentInc_GovPS	Total recurrent funding—Government Per Student	Number	Delete
CE_TotCapExpendPS	Total capital expenditure_ \$ Total Per Student	Number	Delete
D_TotalDeductionsPerStudent	Subtotal_ \$ Per student	Number	Delete
D_TotNetRecurrentIncPS	Total net recurrent income_ \$ Per student	Number	Delete
Stud_Attend_Rate_Year_1_10	Attendance Rate	0–100	Top/Bottom Code & Round

Variable Name	Variable Label	Values	Confidentialisation
Indigenous_Student_Percent	Indigenous_Student_Percent	0–100	Top/Bottom Code & Round
LBOTE_Proportion	LBOTE Proportion	0–100	None
School_ICSEA_Q1	School_ICSEA_Q1	0–100	Delete
School_ICSEA_Q2	School_ICSEA_Q2	0–100	Delete
School_ICSEA_Q3	School_ICSEA_Q3	0–100	Delete
School_ICSEA_Q4	School_ICSEA_Q4	0–100	Delete
Prop_students_to_uni	Proportion of students moving to University	0–100	Top/Bottom Code & Round
Prop_students_to_tafe	Proportion of students moving to TAFE / vocational	0–100	Top/Bottom Code & Round
Prop_students_to_emp	Proportion of students moving to employment destinations	0–100	Top/Bottom Code & Round
y#_R_assessed_percent	Reading (assessed) (Percentage)	0–100	Round
y#_W_assessed_percent	Writing (assessed) (Percentage)	0–100	Round
y#_S_assessed_percent	Spelling (assessed) (Percentage)	0–100	Round
y#_G_assessed_percent	Grammar and Punctuation (assessed) (Percentage)	0–100	Round
y#_N_assessed_percent	Numeracy (assessed) (Percentage)	0–100	Round
y#_R_exempt_percent	Reading (Exempt) (Percentage)	0–100	None
y#_W_exempt_percent	Writing (Exempt) (Percentage)	0–100	None
y#_S_exempt_percent	Spelling (Exempt) (Percentage)	0–100	None
y#_G_exempt_percent	Grammar and Punctuation (Exempt) (Percentage)	0–100	None
y#_N_exempt_percent	Numeracy (Exempt) (Percentage)	0–100	None
y#_R_absent_percent	Reading (absentee) (Percentage)	0–100	None
y#_W_absent_percent	Writing (absentee) (Percentage)	0–100	None
y#_S_absent_percent	Spelling (absentee) (Percentage)	0–100	None
y#_G_absent_percent	Grammar and Punctuation (absentee) (Percentage)	0–100	None
y#_N_absent_percent	Numeracy (absentee) (Percentage)	0–100	None
y#_R_withdr_percent	Reading (withdrawn) (Percentage)	0–100	None
y#_W_withdr_percent	Writing (withdrawn) (Percentage)	0–100	None
y#_S_withdr_percent	Spelling (withdrawn) (Percentage)	0–100	None
y#_G_withdr_percent	Grammar and Punctuation (withdrawn) (Percentage)	0–100	None
y#_N_withdr_percent	Numeracy (withdrawn) (Percentage)	0–100	None
y#_R_mean	Reading mean NAPLAN score	0–1000	Top/Bottom Code & Round
y#_SS_R_mean	Similar Schools Reading Mean	0–1000	Top/Bottom Code & Round
y#_W_mean	Writing mean NAPLAN score	0–1000	Top/Bottom Code & Round

Variable Name	Variable Label	Values	Confidentialisation
y#_SS_W_mean	Similar Schools Writing Mean	0–1000	Top/Bottom Code & Round
y#_S_mean	Spelling mean NAPLAN score	0–1000	Top/Bottom Code & Round
y#_SS_S_mean	Similar Schools Spelling Mean	0–1000	Top/Bottom Code & Round
y#_G_mean	Grammar and Punctuation mean NAPLAN score	0–1000	Top/Bottom Code & Round
y#_SS_G_mean	Similar Schools Grammar and Punctuation Mean	0–1000	Top/Bottom Code & Round
y#_N_mean	Numeracy mean NAPLAN score	0–1000	Top/Bottom Code & Round
y#_SS_N_mean	Similar Schools Numeracy Mean	0–1000	Top/Bottom Code & Round
y#_S_R_mean	NG School Reading mean NAPLAN score	0–1000	Top/Bottom Code & Round
y#_S_N_mean	NG School Numeracy mean NAPLAN score	0–1000	Top/Bottom Code & Round
y#_S_W_mean	NG School Writing mean NAPLAN score	0–1000	Top/Bottom Code & Round
y#_Sim_stud_R_mean	NG Similar students Reading mean NAPLAN score	0–1000	Top/Bottom Code & Round
y#_Sim_stud_N_mean	NG Similar students Numeracy mean NAPLAN score	0–1000	Top/Bottom Code & Round
y#_Sim_stud_W_mean	NG Similar students Writing mean NAPLAN score	0–1000	Top/Bottom Code & Round
y#_Same_start_R_mean	NG Same starting scores Reading mean NAPLAN score	0–1000	Top/Bottom Code & Round
y#_Same_start_N_mean	NG Same starting scores Numeracy mean NAPLAN score	0–1000	Top/Bottom Code & Round
y#_Same_start_W_mean	NG Same starting scores Writing mean NAPLAN score	0–1000	Top/Bottom Code & Round
y#_AS_R_mean	NG All schools Reading mean NAPLAN score	0–1000	None
y#_AS_N_mean	NG All schools Numeracy mean NAPLAN score	0–1000	None
y#_AS_W_mean	NG All schools Writing mean NAPLAN score	0–1000	None
y#_S_R_median	NG School Reading median NAPLAN score	0–1000	Top/Bottom Code & Round
y#_S_N_median	NG School Numeracy median NAPLAN score	0–1000	Top/Bottom Code & Round
y#_S_W_median	NG School Writing median NAPLAN score	0–1000	Top/Bottom Code & Round
y#_Sim_stud_R_med	NG Similar students Reading median NAPLAN score	0–1000	Top/Bottom Code & Round
y#_Sim_stud_N_med	NG Similar students Numeracy median NAPLAN score	0–1000	Top/Bottom Code & Round
y#_Sim_stud_W_med	NG Similar students Writing median NAPLAN score	0–1000	Top/Bottom Code & Round
y#_Same_start_R_med	NG Same starting scores Reading median NAPLAN score	0–1000	Top/Bottom Code & Round

Variable Name	Variable Label	Values	Confidentialisation
y#_Same_start_N_med	NG Same starting scores Numeracy median NAPLAN score	0–1000	Top/Bottom Code & Round
y#_Same_start_W_med	NG Same starting scores Writing median NAPLAN score	0–1000	Top/Bottom Code & Round
y#_AS_R_median	NG All schools Reading median NAPLAN score	0–1000	None
y#_AS_N_median	NG All schools Numeracy median NAPLAN score	0–1000	None
y#_AS_W_median	NG All schools Writing median NAPLAN score	0–1000	None
calendar_year	Calendar Year	2008, 2009, 2010, 2011	None
Geo_Location_Type	Location type	Metropolitan, Provincial, Remote, Very Remote	None
School_Sector_Code	School Sector	G, C, I, which stands for Government, Catholic and Independent	None
School_State	State	NSW, VIC, QLD, SA, WA, TAS, ACT, NT	None
School_Type	School type	Combined, Primary, Secondary, Special	None
y#_Cohort_Range	Cohort Range	3–5, 5–7, 7–9	None
Year_Range	Year Range	String	None

Note: # indicates year level, i.e., 3, 5, 7 or 9.

References

- Australian Curriculum Assessment and Reporting Authority. (2012a). *Guide to understanding Index of Community Socio-educational Advantage (ICSEA)*. Fact Sheet. Sydney: ACARA.
- Australian Curriculum Assessment and Reporting Authority. (2013a). *Guide to understanding the 2013 Index of Community Socio-educational Advantage (ICSEA)*. Fact Sheet. Sydney: ACARA.
- Australian Curriculum Assessment and Reporting Authority. (2011a). *Guide to understanding Index of Community Socio-educational Advantage (ICSEA)*. Fact Sheet. Sydney: ACARA.201
- Australian Curriculum Assessment and Reporting Authority. (2011b). *National Report on Schooling in Australia 2009*. Sydney: ACARA.
- Australian Curriculum Assessment and Reporting Authority. (2012b). *National Report on Schooling in Australia 2010*. Sydney: ACARA.
- Australian Curriculum Assessment and Reporting Authority. (2013b). *National Report on Schooling in Australia 2011*. Sydney: ACARA.
- Australian Curriculum Assessment and Reporting Authority. (2010). *Report on the generation of the 2010 Index of Community Socio-educational Advantage (ICSEA)*. My School Technical Report, March. Sydney: ACARA.
- Australian Curriculum Assessment and Reporting Authority. (2013c). *Report on the generation of the 2012 Index of Community Socio-educational Advantage (ICSEA)*. My School Technical Report, March. Sydney: ACARA.
- Daraganova, G., Edwards, B., & Siphthorp, M. (2013). *Using National Assessment Program Literacy and Numeracy (NAPLAN) data in the Longitudinal Study of Australian Children (LSAC)*. LSAC Technical paper No. 8. Melbourne: Australian Institute of Family Studies.
- Daraganova, G., & Siphthorp, M. (2011). *Wave 4 weights*. LSAC Technical paper No. 9. Melbourne: Australian Institute of Family Studies.
- Edwards, B., Taylor, M., & Fiorini, M. (2011). Who gets the 'gift of time' in Australia? Exploring delayed primary school entry. *Australian Review of Public Affairs*, 10(1), 41–60.
- Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research*, 66(3), 361–396.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London and New York: Routledge.
- Marsh, H. W., & Rowe, K. J. (1996). The effects of single-sex and mixed-sex mathematics classes within a co-educational school: A reanalysis and comment. *Australian Journal of Education*, 40(2), 147–162.
- Rowe, K. J. (1998). Single-sex and mixed-sex classes: The effects of classroom type on student achievement, confidence and participation in mathematics. *Australian Journal of Education*, 32(2), 180–202.