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. 21

Australian Early Development Census (AEDC) data in the Longitudinal Study of Australian Children (LSAC)

Dinusha Bandara, Mark Sipthorp, Fahim Sufi & Galina Daraganova

December 2018

Growing Up in Australia: The Longitudinal Study of Australian Children is conducted in partnership between the Australian Government Department of Social Services, the Australian Institute of Family Studies (AIFS) and the Australian Bureau of Statistics (ABS), with advice provided by a consortium of leading researchers from research institutions and universities throughout Australia growingupinaustralia.gov.au.

Technical paper

The Longitudinal Study of Australian Children: LSAC Technical paper No. 21, Australian Early Development Census (AEDC) data in the Longitudinal Study of Australian Children (LSAC)

Authors: Dinusha Bandara, Mark Sipthorp, Fahim Sufi & Galina Daraganova

For more information, write to:
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

ISBN 978-1-76016-199-6 (Online)
ISBN 978-1-76016-200-9 (PDF)

Contents

List of tables

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. Findings and views expressed in this publication are those of the individual authors and may not reflect those of AIFS, the DSS or the ABS.

The authors wish to acknowledge the support of the DSS, the ABS, the Department of Education and the Social Research Centre, who all worked together to make this data linkage possible.

Thanks, should be offered to Simon Usback at the ABS and Jatender Mohal at AIFS for reviewing data files for validation and confirmation. For useful comments, thanks also go to Constantine Gasser and Dr Diana Warren at AIFS.

For more information, write to:
National Centre for Longitudinal Data
Policy Evidence Branch
Australian Government Department of Social Services
PO Box 7576
Canberra Business Centre ACT

Email: NCLD@dss.gov.au

1. Introduction

This paper aims to describe the linkage and matching between the Longitudinal Study of Australian Children (LSAC) data and the Australian Early Development Census (AEDC) data.

1.1 The Longitudinal Study of Australian Children

Growing Up in Australia: The Longitudinal Study of Australian Children (LSAC) is a national study designed to provide an in-depth understanding of children's development and lifetime wellbeing in Australia's current social, economic and cultural environment, thereby contributing to the evidence base for future policy and program development.

The study is conducted in partnership between the Australian Government 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 from research institutions and universities across Australia.

The study commenced in 2004 with the recruitment of two cohorts: one cohort of 5,107 children aged 0–1 years old (the baby or 'B cohort') and another of 4,983 children aged 4–5 years old (the kindergarten or 'K cohort'), and their families across all states and territories in Australia.

For these analyses, we only used data from the baby or 'B cohort'. The first year of primary school (kindergarten, prep, reception or foundation) for most of the B cohort children coincided with the first AEDC collection, which was in 2009. Children from the B cohort were aged 5–6 years in 2009, providing a valuable opportunity to link to AEDC results.

Table 1: Wave, year and age of LSAC B cohort children

Wave	Year	Age of children
1	2004	0–1 years
2	2006	2–3 years
3	2008	4–5 years
4	2010	6–7 years
5	2012	8–9 years
6	2014	10–11 years
7	2016	12–13 years
8	2018	14–15 years

1.2 Australian Early Development Census

The Australian Early Development Census (AEDC) is a nationwide data collection of early childhood development at the time children commence their first year of primary school (Department of Education and Training, 2018a). The AEDC website provides complete information about the census (www.aedc.gov.au).

The AEDC data provide a snapshot of children's development when they start school to support health, education, community policy, and planning. It measures the five following domains of early child development:

The AEDC is also a useful predictor of child future development and learning, indicating how well early childhood education programs have prepared children for future learning experiences (Department of Education and Training, 2018a).

Formerly known as the Australian Early Development Index (AEDI), AEDC data were first collected in 2009 providing an opportunity to reflect on what the influences may have been for children before arriving at school. Since then, the AEDC data have been collected every third year (2012, 2015 and 2018). The data have been collected from teachers of children in their first year of primary school, using the Australian version of the Early Development Instrument (Department of Education and Training, 2018b). This instrument has been used as a research tool.

	*Physical health and wellbeing* – children's physical readiness for the school day, physical independence and gross and fine motor skills
	*Social competence* – children's overall social competence, responsibility and respect, approach to learning and readiness to explore new things
	*Emotional maturity* – children's pro-social and helping behaviours, and absence of anxious and fearful behaviour, aggressive behaviour and hyperactivity, and inattention
	*Language and cognitive skills (school-based)* – children's basic literacy, interest in literacy, numeracy and memory, advanced literacy and basic numeracy
	*Communication skills and general knowledge* – children's communication skills and general knowledge based on broad developmental competencies and skills

Source: Table 1, Department of Education and Training (2016).

AEDC data are available as Macrodata, Microdata and Microdata for data linkage (Department of Education and Training, 2017).

- Macrodata: Unit record data at an aggregated level in a range of formats to support academic, research or organisational work on policies, planning and service delivery.
- Microdata: Unit record data that contain de-identified and confidential information about individual children. This includes demographic data and AEDC domain scores that have not been anonymised or subjected to statistical disclosure controls that alter unit record information.
- Microdata for data linkage: Unit record data for data linkage projects so that AEDC data can be linked to other datasets specifically for statistical, research and policy purposes.

Microdata for data linkage at child level was used in this technical paper with reference to the Commonwealth Arrangements and complying to principles outlined in the *Guide for Data Integration Projects Involving Commonwealth Data for Statistical and Research Purposes* by the Australian Government (Australian Government National Statistical Service, 2010).

2. LSAC-AEDC data linkage

This section outlines the process that was undertaken to link the LSAC data and the AEDC data. It includes information on obtaining consent from LSAC respondents, the eligible LSAC sample for AEDC data linkage, authorities involved in LSAC-AEDC data linkage, LSAC-AEDC data linkage and the matching process, and the outcome of the LSAC-AEDC data linkage.

2.1 Obtaining consent from the LSAC cohort

At the LSAC Wave 4 data collection (2010), parents of the B cohort children were asked to fill in a consent form allowing access to their study child's AEDC data (Appendix A, page 13). Table 2 (page 3) reports the consent response based on the available sample at Wave 4 and also the total LSAC sample (Wave 1 sample).

At Wave 4 interviews, the majority of LSAC parents returned consent forms for AEDC data linkage (97%, 4,110 children). As a percentage of the total LSAC sample (i.e. those interviewed at Wave 1), this rate was 80%. These forms were then reviewed for their validity.

Table 2: AEDC consent response, B cohort

	Available sample [a]		Total sample [b]	
	n	%	*N*	%
Consent forms returned	4,110	96.9	4,110	80.5
Consent forms not returned	73	1.7	73	1.4
Consent refused during interview	36	0.8	36	0.7
Out of scope – not eligible to receive a consent form	23	0.5	23	0.5
Out of scope – non-participation in Wave 4	–	–	865	16.9
Total	**4,242**	**100.0**	**5,107**	**100.0**

Notes: [a] Available sample refers to families who participated in Wave 4. [b] Total sample refers to Wave 1 sample.
Source: LSAC, B cohort.

Nearly 2% of LSAC parents did not return AEDC consent forms during Wave 4 interviews, and a further 36 parents (less than 1%) refused consent during interviews.

A small number of children were classified as out of scope as they were not eligible to receive a consent form. Reasons for classifying the child as out of scope, included 'special needs child/intellectual disability', 'home-schooled child' and 'distance education child'. Of the total LSAC sample (5,107 who were interviewed at Wave 1), 865 did not participate in Wave 4 interviews and were also considered out of scope.

2.2 Eligible LSAC sample for AEDC data linkage

First criterion: The first eligibility criterion for AEDC linkage to the LSAC sample was that the parents provided valid consent at Wave 4 interviews and the children took part in the LSAC study at the time of linkage (Wave 6).

Table 3 reports the AEDC data linkage consents validity. Initially, 4,110 parents returned the consent forms. However, for consent to be obtained, one of the parents or guardians had to complete the form and sign it in the presence of a witness. If the form was incomplete, it was considered that consent had not been given. Therefore, these forms were reviewed for their validity, and 188 were identified as invalid consents due to invalid forms (where parents filled in the form incorrectly by providing incorrect information or incorrect signatures).

Those who opted out of the LSAC study by the time of the linkage (108 children) were also excluded due to the conditions of the consent agreement. The information sheet provided with the consent forms states that 'If you decide to withdraw from the study or withdraw your consent for the data release, your agreement for the release of the data ceases from the date of your withdrawal' (see Appendix B, page 14). Therefore, 3,814 children were identified as satisfying the first criterion.

Table 3: AEDC data linkage consent validity

	n	(%)
Invalid consents: form filled in incorrectly	188	4.6
Invalid consents: opt-outs of the study by the time of linkage	108	2.6
Valid consents at the time of linkage	3,814	92.8
Total forms returned	4,110	100.0

Source: LSAC, B cohort

Second criterion: The second eligibility criterion (independent of the first criterion) was that the LSAC study child was in their first year of primary school in 2009. LSAC data showed that 3,074 LSAC children started school in 2009. However, 24 of these children repeated the school year. Therefore, only 3,050 LSAC children were in their first year of primary school in 2009.

Eligible LSAC sample: Overall, 2,765 had valid consents at the time of linkage and were in their first year of primary school in 2009 (satisfying both criteria).

2.3 Authorities involved in LSAC-AEDC data linkage

The AEDC data, identified as needing linkage, were held in one central location, The Social Research Centre (SRC). The LSAC data were managed by the Australian Government Department of Social Services (DSS), the Australian Institute of Family Studies (AIFS) and the Australian Bureau of Statistics (ABS). Therefore, this was established through a Letter of Agreement between the Department of Education and Training, the SRC, the ABS, DSS and AIFS.[1]

The agencies required to complete the linkage process were:

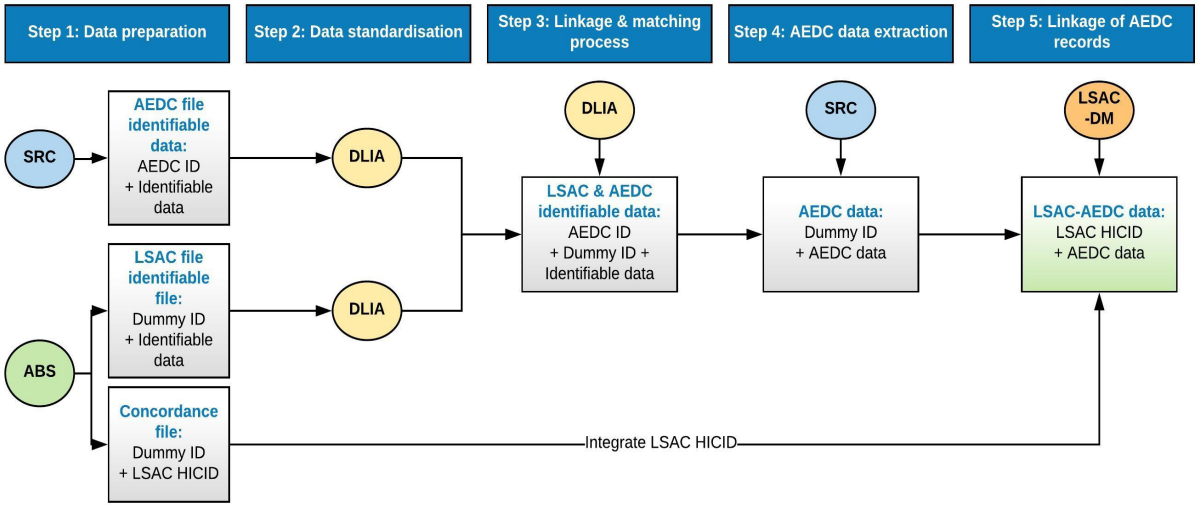
1. Social Research Centre (SRC)
2. Australian Bureau of Statistics (ABS)
3. AIFS' Data Linkage and Integrating Authority (DLIA)
4. AIFS' LSAC Data Management team (LSAC-DM).

These agencies completed the LSAC-AEDC data linkage for children who were part of the eligible sample.

2.4 LSAC-AEDC data linkage and matching process

As illustrated in Figure 1, the LSAC-AEDC data linkage and matching process included five steps: data preparation; data standardisation; linkage and matching process; AEDC data extraction; and linkage of AEDC records.

Figure 1: LSAC-AEDC data linkage and matching process



1 AIFS is an accredited Commonwealth Data Linkage and Integrating Authority and has met stringent criteria covering project governance, data linkage capability and information management. Maintained as a distinct functional unit within AIFS, the DLIA handles linkage information in a secure manner. This is achieved through the key protective protocol that underlies the DLIA's data management: the *separation principle* (Kelman, Bass, & Holman, 2002). AIFS' LSAC Data Management team (LSAC-DM) performs data management tasks independent of the DLIA with no crossover.

Step 1: Data preparation

AEDC file with identifiable data

The SRC compiled a file with identifiable data for all children who started school in 2009 for whom a teacher completed a report for the first year of school. The file included AEDC ID, and identification variables for all 557,518 records available at the time of linkage (including all 261,147 records from 2009 data collection). The identifiable information provided in the file included:

- first name
- surname
- gender
- date of birth
- school name
- school address.

The file was hosted on the SRC's Secure File Transfer Protocol site and provided to the DLIA with the credentials to download the file.

LSAC file with identifiable data

ABS compiled a list of 3,814 LSAC children (who satisfied the first eligible criterion), with a Dummy ID and aforementioned identifying variables. It was decided to compile the list based on the first eligible criterion, rather than both criteria, so that if the start of the first year of school was missed in LSAC interview data, it would be identified from AEDC data. The ABS provided the password-protected encrypted file and supplied the password separately to the DLIA.

ABS concordance file

The ABS also compiled the concordance file containing the Dummy ID and the HICID (the LSAC unique identifier).

Step 2: Data standardisation

Data standardisation procedures undertaken to link AEDC and LSAC data included the following:

1. Resolving issues raised when matching names due to typographical errors by:
   - changing the case on all the strings
   - removing punctuation
   - removing prefixes and suffixes
   - ignoring middle initials
   - looking for transpositions in words and initials
   - identifying nicknames.
2. Resolving issues raised in the presence of null values or empty strings by:
   - deleting consecutive spaces
   - trimming trailing or leading spaces.
3. Resolving issues raised in content dissimilarity by:
   - resorting to manually choosing the trusted source between the two files, when only one identifier is dissimilar in comparison (while other identifiers are identical).
4. Resolving issues that arise in content integrity by:
   - using checksums (i.e. a method which validates an ID using a mathematical algorithm) to find invalid unique identifiers in the AEDC ID to avoid duplication of the Student ID within AEDC data
   - generating a composite ID using both year and student ID for uniquely identifying an AEDC record.
5. Resolving issues raised in date formats for date of birth by:
   - unifying date formats
   - checking for transposition in dates
   - finding automatically filled dates or dates too far in the future/past.

Step 3: Linkage and matching process

Subsequent to data standardisation procedures, the DLIA followed a rigorous rule-based methodology (fuzzy logic and rule-based classifier) using similarity scores and confidence measures to perform the linkage and matching process of the LSAC file and the ABS file containing identifiable data or records.

- *Similarity scores*: the parameter used to describe the average similarity between values in the input and reference columns.
- *Confidence measure*: the parameter used to describe the quality of the match.

Table 4 provides an aggregated summary of the identifiable data linkage process in terms of probability of similarity (a match). On average, the first name field was 97% similar, the last name field was 99% similar, gender was 99% similar, and the school name and address were 96% similar to the linked records. In summary, the linkage and matching process of identifiable data was completed with a very high level of assurance and confidence (97% and 92% respectively).

Table 4: Aggregated summary of linkage of identifiable data

	Average	*Standard deviation*
First name similarity	0.9745	0.0937
Last name similarity	0.9907	0.0564
Gender similarity	0.9880	0.1079
School name and address	0.9639	0.0998
Overall similarity	0.9681	0.0674
Confidence	0.9201	0.1750

Using optimal rules (complex rules that can be used to classify pairs of records as matches or as non-matches), allowing the capture of as many links as possible with a high level of accuracy and confidence (Christen & Goiser, 2007), 2,418 AEDC records were classified as 'sure matched records'. From the remaining AEDC records, 194,976 were classified as 'possible match records'. A further 360,124 were classified as 'no match records', which included those outside of the 2009 AEDC data collection.

The linkage and matching process of identifiable data revealed 2,418 'sure matched records'. However, further review of these identifiable data or records identified one false positive, reducing 'sure matched records' to 2,417. During the 'possible match records' review, 44 records were identified as matched records (totaling the match records to 2,461) and two duplicates were also identified resulting in 2,459 as the final linked identifiable data.

Step 4: AEDC data extraction

The DLIA uploaded the linked file of these data records to the SRC's File Transfer Protocol. This file contained the AEDC ID supplied by SRC, the Dummy ID supplied by the ABS and identifiable information. The SRC then extracted the AEDC records of the students for whom the AEDC ID was provided. SRC then removed AEDC ID and identifiable information.

From these 2,459 children, SRC identified AEDC data for 2,264 children. However, the LSAC sample was linked to the AEDC data based on the first eligibility criterion, and therefore SRC identified a further 195 records, where parents missed or provided the incorrect start of the first year of primary school during the LSAC interviews.

Step 5: Linkage of AEDC records

To complete the data linkage a number of further steps were taken by LSAC-DM. These steps had a number of security measures in place and conformed to the separation principle, that is mandatory in all data integration projects within AIFS. Steps included:

1. The SRC sent the AEDC data to LSAC-DM. This only contained the Dummy ID and the AEDC survey variables.
2. The ABS then sent the concordance file to the LSAC-DM that contains the Dummy ID and the HICID (the LSAC unique identifier).
3. The LSAC-DM replaced the Dummy ID with the HICID and added the data to the LSAC data package.

2.5 Outcome of LSAC-AEDC data linkage

In the 2009 AEDC data collection, information was collected on 261,147 Australian children (97.5% of the estimated five-year-old population) in their first year of primary school between 1 May and 31 July 2009 (Centre for Community Child Health and Telethon Institute for Child Health Research, 2009).

The eligible LSAC sample for LSAC-AEDC data linkage was 2,765 children whose parents provided valid consent at Wave 4 interviews, and who participated in the study at the time of linkage and started the first year of primary school in 2009. From these 2,765 children, 2,459 were linked to AEDC data, a match rate of 89%.

Overall, the LSAC-AEDC data integration and linkage process successfully linked **2,459** children from the eligible LSAC cohort, and is referred to as the LSAC-AEDC cohort throughout this technical paper.

3. Children in the LSAC-AEDC cohort and in the national AEDC

This section outlines the alignment of the LSAC–AEDC cohort to the national AEDC and briefly discusses the demographic profile of the children, and their experience of child care and early education before entering primary school.

3.1 Demographic profile of children in LSAC-AEDC and in the national AEDC

In 2009, the national AEDC represented 97.5% of the estimated five-year-old population in Australia. Overall, 261,147 children were included in the data collection, where 15,522 teachers and 7,422 schools contributed to the AEDC results nationally (Centre for Community Child Health and Telethon Institute for Child Health Research, 2009). Table 5 (page 8) compares the demographic profile at a national level to the demographic profile of the LSAC-AEDC cohort.

Overall, when child age, sex, Indigenous status, language background, socio-economic status and state/territory were compared, only small differences were evident when comparing the LSAC-AEDC cohort to the national AEDC data. For example, all LSAC B cohort children were born in Australia. Therefore, the proportion of children with English as a primary language was slightly higher among the LSAC-AEDC cohort than in the national AEDC cohort (89.7% vs 87.2%). The difference in age reflects that LSAC is a birth cohort not a year level cohort and LSAC children who were enrolled in school earlier or later than their eligible age would have enrolled in school in 2008 or 2010, respectively. The data for these calendar years are not available as there were no AEDC data collections in 2008 and 2010.

3.2 Children's experiences in the year before entering primary school

As part of the national AEDC Checklists, teachers were asked to record children's experiences in the year before entering primary school (Centre for Community Child Health and Telethon Institute for Child Health Research, 2009; Department of Education and Training, 2009). The teachers were able to nominate more than one form of care or education type for a child.

Overall, 65% of children were reported to have been in preschool or kindergarten, 24% in a day care centre with a preschool or kindergarten program, 8.4% in a day care centre without a preschool or kindergarten program, 14.7% in a day care centre but unsure about a preschool or kindergarten program and 8.3% in family day care. In LSAC, the study child's main carer (usually their mother) was asked a series of questions about the study child's attendance at early education and care settings. When study children were aged 4–5, parents reported the child's ECEC arrangements at the time of interview. Parents also provided information retrospectively, based on what the children were doing in the previous year. At four years of age, most children for whom AEDC data were linked (83%) attended a preschool program.

Table 5: Demographic profile of children in LSAC-AEDC and national AEDC

	LSAC-AEDC [a] (*N* = 2,459) % [95% CI]	National-AEDC [b] (*N* = 261,147) %
Mean age	5 years and 6 months	5 years and 7 months
Age group [c]		
<4 years	0	4.0
>= 5 and <6 years	92.5 [91.5–93.6]	78.9
>=6 and <7 years	7.5 [6.4-–8.5]	17.0
>=7 years	0	0.1
Sex		
Male children	52.7 [50.8–54.7]	51.3
Female children	47.3 [45.3–49.2]	48.7
Indigenous status		
Indigenous children	3.4 [2.7–4.1]	4.8
Non-Indigenous children	96.6 [95.9–97.3]	95.2
Language		
Children with English as a second language	10.3 [9.1–11.5]	12.8
Children with English as a primary language	89.7 [88.5–90.9]	87.2
Socio-economic status		
Quintile 1 (most disadvantaged)	19.4 [17.9–21.0]	20.6
Quintile 2	20.0 [18.4–21.6]	19.1
Quintile 3	20.7 [19.1–22.4]	19.3
Quintile 4	21.0 [19.4–22.6]	19.8
Quintile 5 (least disadvantaged)	18.8 [17.3–20.4]	21.1
State/Territory		
New South Wales	33.0 [31.2–34.9]	33.3
Victoria	25.8 [24.1–27.5]	23.5
Queensland	19.2 [17.7–20.8]	21.2
Western Australia	6.1 [5.1–7.0]	10.6
South Australia	8.4 [7.3–9.5]	6.2
Tasmania	3.4 [2.7–4.1]	2.3
Australian Capital Territory	1.1 [0.7–1.5]	1.8
Northern Territory	3.1 [2.4–3.8]	1.2

Note: CI = Confidence Interval.

Source: [a] LSAC, B cohort, [b] Department of Education and Training (2016). [c] Department of Education and Training (2009).

4. Early childhood development results in the LSAC-AEDC and the national AEDC

In every AEDC checklist (completed by a child's teacher), children receive a score for each domain. Each checklist consists of more than 100 questions measuring the five developmental domains. The scores range from 0 (the lowest score) to 10 (the highest score) (Centre for Community Child Health and Telethon Institute for Child Health Research, 2009).

In both LSAC-AEDC cohort children and national children, the majority of children were developmentally on track for each of the five AEDC domains:

1. physical health and wellbeing
2. social competence
3. emotional maturity
4. language and cognitive skills (school-based)
5. communication skills and general knowledge.

4.1 Development vulnerability of children in the LSAC-AEDC and in the national AEDC

To determine which children fell into the 'developmentally vulnerable' group, AEDC cut-offs were set for each domain on the basis of the National-AEDC population. Children who fell below the 10th percentile of the National-AEDC population were classified as 'developmentally vulnerable' (Centre for Community Child Health and Telethon Institute for Child Health Research, 2009).

Derived information on children who fell below the 10th percentile for each domain was included in the AEDC data linked to the LSAC-AEDC cohort (Department of Education and Training, 2016). The cut-offs were based on 2009 AEDC census results. Summarising this information, Table 6 reports the developmental vulnerability in each domain for the LSAC-AEDC cohort and the national AEDC.

Table 6: Developmental vulnerability of children

Domain	LSAC-AEDC [a] % [95% CI] (*N**)	National-AEDC [b] % (*N**)
Physical health and wellbeing	6.3 [5.3–7.3] (2,317)	9.3 (247,232)
Social competence	6.7 [5.6–7.7] (2,316)	9.5 (247,189)
Emotional maturity	7.0 [6.0–8.1] (2,305)	8.9 (246,197)
Language and cognitive skills (school-based)	4.6 [3.7–5.4] (2,313)	8.9 (246,810)
Communication skills and general knowledge	5.0 [4.1–5.9] (2,316)	9.2 (247,212)

Notes: CI = Confidence Interval. * Total children with valid scores at school entry.
Source: [a] LSAC, B cohort, [b] Department of Education and Training (2016).

In all domains, the percentage of children with developmental vulnerability in the LSAC-AEDC was less than those in the national AEDC. For LSAC-AEDC children, the highest developmental vulnerability percentage was seen in the emotional maturity domain. Characteristics of children developmentally vulnerable in this domain (Centre for Community Child Health and Telethon Institute for Child Health Research, 2009) included:

- *pro-social and helping behaviour*: They never or almost never show most of the helping behaviours, including helping someone hurt, sick or upset, offering to help spontaneously, and inviting others to join in.
- *anxious and fearful behaviour*: Often show most of the anxious behaviours; they could be worried, unhappy, nervous, sad or excessively shy, indecisive; and they can be upset when left at school.
- *aggressive behaviour*: Often show most of the aggressive behaviours; they get into physical fights, kick or bite others, take other people's things, are disobedient or have temper tantrums.

- hyperactivity and inattention: Often show most of the hyperactive behaviours; they could be restless, distractible, impulsive; they fidget and have difficulty settling to activities.

Among the LSAC-AEDC children, the lowest percentage of developmental vulnerability was seen in the language and cognitive skills (school-based) domain. Characteristics of children developmentally vulnerable in this domain (Centre for Community Child Health and Telethon Institute for Child Health Research, 2009) included:

- *basic literacy*: Do not have most of the basic literacy skills; they have problems with identifying letters or attaching sounds to them, rhyming, may not know the writing directions and how to write their own name.
- *interest in literacy/numeracy and memory*: They may not show interest in books and reading, or maths and number games, or both, and may have difficulty remembering things.
- *advanced literacy*: Have only up to one of the advanced literacy skills; they cannot read or write simple words or sentences, and rarely write voluntarily.
- *basic numeracy*: Have marked difficulty with numbers, they cannot count, compare or recognise numbers, may not be able to name all the shapes and may have difficulty with time concepts.

Characteristics of children developmentally vulnerable in the social competence domain (Centre for Community Child Health and Telethon Institute for Child Health Research, 2009) included:

- *overall social competence*: They have average to poor overall social skills, low self-confidence and are rarely able to play with various children and interact cooperatively.
- *responsibility and respect*: They only sometimes or never accept responsibility for actions, show respect for others and for property, demonstrate self-control, and are rarely able to follow rules and take care of materials.
- *approaches to learning*: Only sometimes or never work neatly and independently, they are rarely able to solve problems and follow class routines, and do not easily adjust to changes in routines.
- *readiness to explore new things*: They only sometimes or never show curiosity about the world, and are rarely eager to explore new books, toys or unfamiliar objects and games.

Characteristics of children developmentally vulnerable in the physical health and wellbeing domain (Centre for Community Child Health and Telethon Institute for Child Health Research, 2009) included:

- *physical readiness for school day*: They have, at least sometimes, experienced coming unprepared for school by being dressed inappropriately, or coming to school hungry or tired.
- *physical dependence*: Range from those who have not developed one of the three skills (independence, handedness, co-ordination), to those who have not developed any of the skills.
- *gross and fine motor skills*: Range from those who have an average ability to perform skills requiring gross and fine motor competence and good or average overall energy levels, to those who have poor gross and fine motor skills, poor overall energy levels and physical skills.

Characteristics of children developmentally vulnerable in the communication skills and general knowledge domain (Centre for Community Child Health and Telethon Institute for Child Health Research, 2009) included:

- *communication skills and general knowledge*: Range from being average to very poor in effective communication, they may have difficulty in participating in games involving the use of language, may be difficult to understand and/or have difficulty in understanding others and may show little general knowledge.

Nationally, 23.6% of children were developmentally vulnerable on one or more of the AEDC domain/s (Centre for Community Child Health and Telethon Institute for Child Health Research, 2009), while only 17% (95% CI of 15.4–18.5) of LSAC-AEDC children were considered to be vulnerable in at least one aspect of their development. This suggests that children in the LSAC-AEDC linked cohort have slightly stronger early developmental outcomes than children across the general Australian population.

A comparison of the AEDC-LSAC with the Wave 3 LSAC B cohort, and further analysis of the AEDC data by Warren, Daraganova, and O'Connor (2018) shows that there is an association between preschool experiences and children's development at school entry, as measured by the AEDC, specifically in developmental domains closely related to learning: language, cognition, communication and general knowledge.

4.2 AEDC-LSAC sample and weights

The Longitudinal Study of Australian Children is based on a two-stage clustered sample design, using postcodes as primary sample units. A variety of weights have been used in LSAC. At each wave, sampling and population weights have been calculated. Sampling weights sum the number of children in the sample and show how representative each child is in comparison to other children in the sample. Population weights sum the number of children in the population, and when generating weighted estimates, conceptually show the number of children in the population represented by each child in the sample. Cross-sectional weights adjust the sample from the current wave (e.g. Wave 3) to represent the population at the time of selection. However, longitudinal weights adjust the sample from the current wave to represent the sample that has responded to all waves of LSAC, up until a particular time point (e.g. Waves 1, 2 and 3).

While LSAC weights have been developed, they were not used for comparison of the AEDC-LSAC sample to the national AEDC sample. The AEDC-LSAC sample only includes children who started primary school in 2009 when the AEDC data collection took part. Therefore, it did not include the entire B cohort of LSAC; that is, B cohort children who enrolled in school in 2008 and 2010 and were substantially younger or older at time of school entry compared to B cohort children who enrolled in school in 2009. Also, AEDC data collection took place in 2009 (between LSAC waves), while LSAC weights were generated to account for attrition bias in LSAC data collections that took place in 2008 and 2010. Therefore, caution should be given when analysing AEDC data using LSAC weights.

5. Conclusion

This technical paper has described the process of AEDC data linkage to LSAC data. This included a description of the process of obtaining consent from LSAC respondents, the eligible LSAC sample for AEDC data linkage, the authorities involved in LSAC-AEDC data linkage, LSAC-AEDC data linkage and the matching process (which included five steps: data preparation, data standardisation, a linkage and matching process, AEDC data extraction and linkage of AEDC records), and the outcome of the LSAC-AEDC data linkage. The outcome of the data integration and linkage process was the linked LSAC-AEDC cohort of 2,459 children.

Further, key characteristics of the LSAC-AEDC cohort have been described and compared for the alignment of the LSAC-AEDC cohort to the national AEDC. In both LSAC-AEDC cohort children and national children, the majority of children were developmentally on track for each of the five AEDC domains.

In all domains, children in the LSAC-AEDC showed less developmental vulnerability than the national AEDC. Nationally, 23.6% of children were developmentally vulnerable on one or more of the AEDC domain/s in 2009. In LSAC-AEDC children, 17% were considered to be vulnerable in at least one aspect of their development when they started school in 2009.

In summary, the LSAC-AEDC cohort of **2,459** children is a valuable data resource, with the potential to answer many important and policy-relevant questions. Data users should be aware of the slightly lower levels of vulnerability of children in the LSAC-AEDC cohort, compared to national AEDC levels.

6. References

Australian Government National Statistical Service. (2010). *A guide for data integration projects involving Commonwealth data for statistical and research purposes*. Retrieved from statistical-data-integration.govspace.gov.au/

Centre for Community Child Health and Telethon Institute for Child Health Research. (2009). *A Snapshot of Early Childhood Development in Australia – AEDI National Report 2009*. Retrieved from www.aedc.gov.au/resources/detail/national-report-2009

Christen, P., & Goiser, K. (2007). *Quality and complexity measures for data linkage and deduplication. studies in computational intelligence*. 43. 10.1007/978-3-540-44918-8_6

Department of Education and Training. (2009). *AEDI checklist and response criteria*. Retrieved from www.aedc.gov.au/resources/detail/2009-early-development-instrument-questions

Department of Education and Training. (2016). A snapshot of early childhood development in Australia – Australian Early Development Census national report 2015. Retrieved from www.aedc.gov.au/resources/detail/2015-aedc-national-report

Department of Education and Training. (2016). Australian Early Development Census (AEDC) data dictionary. Retrieved from www.aedc.gov.au/resources/detail/aedc-data-dictionary-2017

Department of Education and Training. (2017). Australian Early Development Census (AEDC) data guidelines 2017. Retrieved from www.aedc.gov.au/resources/detail/aedc-data-guidelines-2017

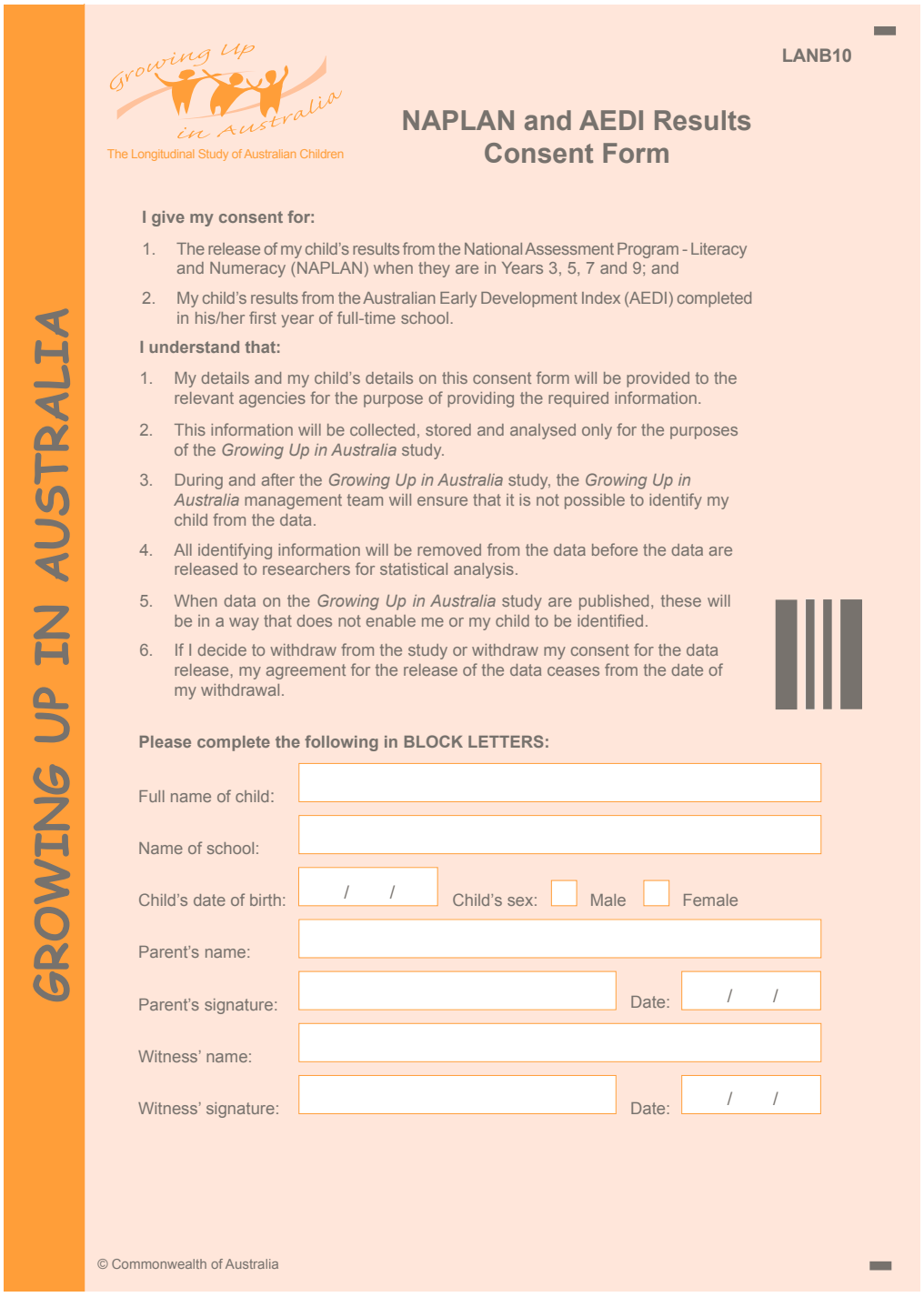
Department of Education and Training. (2018a). Australian Early Development Census. Retrieved from www.aedc.gov.au

Department of Education and Training. (2018b). Data. Retrieved from www.aedc.gov.au/data

Kelman, C. W., Bass, A. J., & Holman, C. D. J. (2002). Research use of linked health data – a best practice protocol. *Aust N Z J Public Health*, *26*(3), 251–255.

Warren, D., Daraganova, G., & O'Connor, M. (2018, forthcoming). *Preschool programs as an opportunity to promote healthy development as children start school*. In preparation by the Australian Institute of Family Studies.

Appendix A



GROWING UP IN AUSTRALIA

Growing Up in Australia
The Longitudinal Study of Australian Children

LANB10

NAPLAN and AEDI Results Consent Form

I give my consent for:

1. The release of my child's results from the National Assessment Program - Literacy and Numeracy (NAPLAN) when they are in Years 3, 5, 7 and 9; and
2. My child's results from the Australian Early Development Index (AEDI) completed in his/her first year of full-time school.

I understand that:

1. My details and my child's details on this consent form will be provided to the relevant agencies for the purpose of providing the required information.
2. This information will be collected, stored and analysed only for the purposes of the *Growing Up in Australia* study.
3. During and after the *Growing Up in Australia* study, the *Growing Up in Australia* management team will ensure that it is not possible to identify my child from the data.
4. All identifying information will be removed from the data before the data are released to researchers for statistical analysis.
5. When data on the *Growing Up in Australia* study are published, these will be in a way that does not enable me or my child to be identified.
6. If I decide to withdraw from the study or withdraw my consent for the data release, my agreement for the release of the data ceases from the date of my withdrawal.

Please complete the following in BLOCK LETTERS:

Full name of child:

Name of school:

Child's date of birth: / / Child's sex: Male Female

Parent's name:

Parent's signature: Date: / /

Witness' name:

Witness' signature: Date: / /

Appendix B

LANB10

The Longitudinal Study of Australian Children

NAPLAN and AEDI Results Information Sheet

Literacy and numeracy assessments are conducted with students across Australia in Years 3, 5, 7 and 9. This is known as the National Assessment Program - Literacy and Numeracy (NAPLAN).

The Australian Early Development Index (AEDI) is a measure of children's development as they enter school. Based on scores from a teacher completed checklist, the AEDI measures areas of early childhood development.

We are seeking your signed consent for the release of results from NAPLAN and AEDI for your *Growing Up in Australia* study child until the end of the study. If you agree, this information will be obtained from the relevant education agency in your state.

The inclusion of this information, along with the other information we are gathering in the study, will allow us to gain further insight into your child's learning and development.

With regard to the release of data:

- Your details and your child's details, as given on the consent form, will be provided to the relevant education agency for the purpose of providing the required information.
- This information will be collected, stored and analysed only for the purposes of the *Growing Up in Australia* study.
- During and after the *Growing Up in Australia* study, the *Growing Up in Australia* management team will ensure that it is not possible to identify your child from the data.
- All identifying information will be removed from the data before the data are released to researchers for statistical analysis.
- Data from the *Growing Up in Australia* study are published in a way that does not enable you or your child to be identified.
- If you decide to withdraw from the study or withdraw your consent for the data release, your agreement for the release of the data ceases from the date of your withdrawal.

You will receive summary results of your child's NAPLAN assessments from your child's school. If you would like further information on these, please contact your child's school or visit the website http://www.naplan.edu.au/

For more information on AEDI visit the website www.rch.org.au/australianedi

If at any stage you have queries about the release of your child's results to *Growing Up in Australia*, please call on 1800 005 508 freecall (excluding mobile phones).