The Longitudinal Study of Australian Children

growingupinaustralia.gov.au

THE LONGITUDINAL STUDY OF AUSTRALIAN CHILDREN:
AN AUSTRALIAN GOVERNMENT INITIATIVE

Data issues

Waves 1 to 9C2

June 2022

Australian Government
Australian Institute of Family Studies

Australian Bureau of Statistics

Australian Government
Department of Social Services

Growing Up in Australia: The Longitudinal Study of Australian Children 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 throughout Australia.

The Release 9C2 data files were prepared by the ABS and AIFS data processing teams and reviewed by DSS. The current version of the LSAC data issues paper was updated by ABS.

Readers wishing to refer to this document should cite the following:
Australian Bureau of Statistics. (2022). Growing Up in Australia: *The Longitudinal Study of Australian Children – Data Issues Waves 1 to 9C2. June 2022*. Melbourne: Australian Institute of Family Studies. doi:10.26193/QR4L6Q

The authors wish to acknowledge:

- DSS and AIFS for their support in providing feedback.
- Former ABS Data Processing team members for their contribution to earlier versions of the data issues paper.

Australian Institute of Family Studies
Level 4, 40 City Road, Southbank VIC 3006 Australia
Phone: (03) 9214 7888 Web: aifs.gov.au

Cover photo: © gettyimages/Cecilie_Arcurs

Contents

List of tables

List of figures

1 Introduction

This paper provides a summary of data-related issues that have emerged over the life of *Growing up in Australia*: The Longitudinal Study of Australian Children (LSAC). The chapters were initially published on the LSAC website as a series of Issues Papers designed to assist users of the LSAC data as they undertake research and analysis of the LSAC datasets.

The paper is to be used in conjunction with the Data User Guide available on the LSAC website.

2 Cleaning of time use diary data

2.1 Background

The LSAC time use diary (TUD) is a diary consisting of 96 15-minute time intervals or bubbles with pre-coded activity (e.g. sleeping, eating, bathing) and context (e.g. where they were and who they were with) information. Parents are asked to mark which of the pre-coded activities were done during each of the 96 time intervals. The diary begins at 4 am. Time interval 1 is from 4 am to 4:14 am, time interval 2 is from 4:15 am to 4:29 am, etc. For the B cohort at Wave 1 there were 22 pre-coded activities, five context locations and seven 'who else was present' context options. Additionally, the diarists were asked whether they had paid for the activity that the child was doing. For the Wave 1 B cohort the total matrix size was 3,360, consisting of the 35 activities and context descriptors by 96 time intervals. The Wave 1 K cohort had 26 pre-coded activities. Otherwise, the diary was the same as for the B cohort. The matrix size for the K cohort was 3,744.

The data entry used scanning technology. For Wave 1, few checks were made at the time of data entry and subsequently it has been found that the scanner was sensitive to rub outs and other marks that appeared in the bubbles on the paper files. This resulted in false data (false positives) that exists in the electronic data files but does not exist on the paper files.

For Wave 2 various procedures were implemented to ensure that these problems did not recur. These procedures involved changes to the data capture and data validation stages.

To reduce problems associated with capturing the data, changes were made to the scanner settings. Through the Intelligent Forms Processing (IFP) system it is possible to define the minimum character/mark size that will be registered by the system. Testing of TUD capture confirmed that oversensitivity of scanning equipment can produce a high rate of false positive responses on the TUD. Following iterative testing of LSAC dress rehearsal (DR) TUD forms, it was determined that the character size for the TUD 'bubbles' should be increased from 2 x 2 pixels to 2 x 5 pixels. Testing showed that this setting allowed the IFP system to disregard very small specs of dust, etc. (thereby greatly reducing false positives) without any impact on the false negative rate.

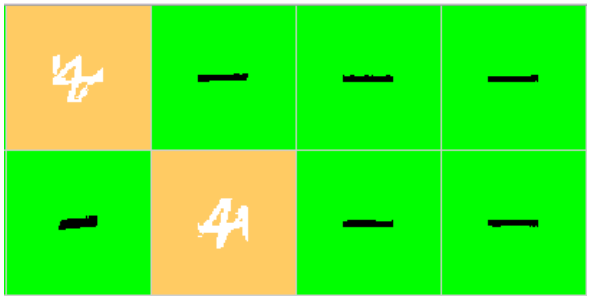
A second setting that impacts on the registration of marks is the size of field that is 'scanned' for a character/mark. Often this field size is expanded slightly beyond the expected capture area to allow marks falling slightly outside the response box to be registered. However, in the case of the TUD, the extremely close arrangement of the response bubbles meant that such an expansion led to false positives from slight (and unintended) continuation of marks beyond one response bubble but not quite into the subsequent one. For this reason, horizontal margins for the capture area have been strictly limited to the intended response area (i.e. the border of the response bubble). However, vertical (top/bottom) margins of 6 pixels outside the response bubble have been retained to ensure that marks made slightly above/below a response bubble are captured.

Following capture, forms are forwarded for inspection and repair by a trained operator. The process outlined below is performed on all TUD forms, with the majority expected to contain at least one response mark that will need to be investigated.

The first repair process conducted on scanned forms is the on-screen inspection of mosaics of scanned response marks known as carpets. Carpets display images of all marks from the same form that have been recognised by the system. Depending on the system's confidence in the validity of a particular response, the mark will be displayed in a green, yellow or red shaded box. At this stage the operator is able to confirm or correct a response or, alternatively, select responses for further investigation through the form process (see below).

In the example below (Figure 1) the carpet displays images of eight response marks, all of which were confidently identified by the system as valid responses. In two cases the respondent has clearly attempted to correct a response by crossing out the original mark. While the Optical Mark Reader (OMR) scanner is unable to distinguish these responses from the valid responses, the analysis of carpets gives the operator an opportunity to correct the data, thereby avoiding potential false positives. Operators conducting repair of LSAC forms are trained to examine carpets for these types of responses.

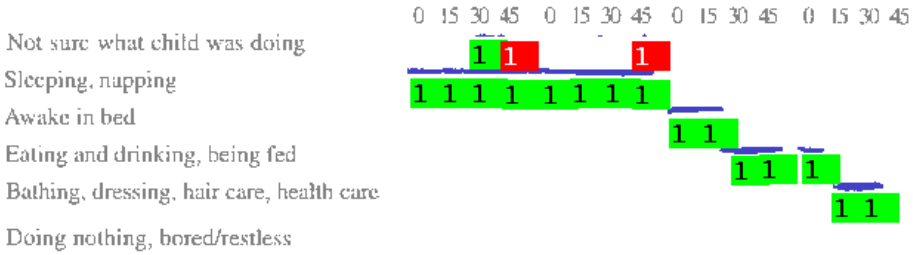
Figure 1: Example of 'carpets' from time use diary scanning



Forms containing at least one mark queried by the system or the operator progress to the forms stage of repair. At this stage the operator is able to see the queried response in the context of the form and other responses.

In the test example below (Figure 2) the operator has queried the two marks highlighted in red during the earlier examination of carpets for this form. The ability to further examine these marks then enables the operator to make a more informed decision regarding its validity.

Figure 2: Example of in-context diary data display



The processes outlined above have been developed as a result of the experience of the Australian Bureau of Statistics (ABS) in the capture of LSAC TUD forms during the Wave 2 dress rehearsal and more recent testing of final forms. It should be noted, however, that these processes do not address data quality issues associated with respondent error nor will they fully overcome capture difficulties associated with formatting features of the TUD such as the extremely close arrangement of response bubbles and the sheer volume of information recorded on each page.

The rest of this section reports on the extent of the Wave 1 false positives and provides a description of attempts to electronically remove the false positives, as well as other measures to improve data quality. A number of strategies were used to recode the data, working off the premise that most implausible data combinations are likely to be false positives. For example, if it is late in the evening or early in the morning, a child is likely to be either in bed or asleep, not simultaneously sleeping and walking (not for extended periods of time at least). The rules for the electronic recodes are outlined in more detail below. These corrections were applied only to the early Wave TUD data.[1]

In recoding the data it is important that the amount of real data being incorrectly removed (false negatives) is minimised. It is expected that this incorrect coding was most likely to occur when there were transitions between activities. In order to protect against this, sequences of events were often considered. That is, comparisons were often made with the preceding and following time interval. However, it must be realised that from time to time diarists unintentionally provide information on implausible events.

1 Note that at the time of the release of Wave 2 data the Wave 2 TUDs also had these processes applied to them in order to maximise consistency. For the release of the Wave 3 data this decision was reversed. It was felt that the corrections excluded some combinations of activities that were unlikely but possible, particularly as the children became older (e.g. sleeping outside).

In addition to the recoding of false positives, other data cleaning strategies or imputations are employed. These recodes are potentially important, as it is common in time use analysis to exclude data with more than 90 minutes of missing data. Thus, if the number of missing bubbles can be minimised, less data is lost. These imputations are performed on the data from all three waves.

2.2 Assessing the extent of the false positives

In order to estimate the extent of the false positives and to ascertain whether corrections could be made electronically, a random sample was drawn from both the B cohort (*n* = 51) and the K cohort (*n* = 49). One diary was excluded from the B cohort as the diary was returned blank, and two diaries were excluded from the K cohort, one because only one activity was given for the whole diary and the other because it did not match the electronic file at all. These forms were manually checked against the electronic records so that false positives were identified and given a unique code on the electronic record. These files are known throughout this report as the 'corrected files'. The files of these cases before they are checked are known as the 'original files'.

K cohort, Wave 1

A summary of false positives by data type for the K cohort is presented in Table 1. Over the random sample of 47 children and over the 96 time intervals, there were 16,248 positive responses in the original file. This file had 882 'extra' bubbles or units of data than was provided in the 'corrected file'. Thus, the false positive rate is 6% (876/(876 + 15,366)*100). It should be noted that there was one diary that had a false positive rate of 30%, while the next highest figure was 14%. If these cases were to be excluded, the false positive rate would drop to 5%.

For the K cohort, the highest aggregate false positive rate was for the 'with whom' context data (7%), while for the activity data and 'where' context data, the false positive rate was around 5%. The light diary also asked whether someone was paid for the activity. There was only one false positive associated with this data.

The general trend was that the more real data there was the greater the number of false positives. This is not surprising given that much of the false positive data was due to rub-outs.

Table 1: Summary of false positives based on comparing original and corrected file of 47 cases (K cohort, Wave 1)

Source	True positives	False positives	False positive rate (%)
Total	**15,366**	**876**	**5.7**
Activity	5,241	274	5.2
Where	3,738	195	5.2
Who	6,231	406	6.5
Paid	156	1	0.6

B cohort, Wave 1

A summary of false positives by data type for the B cohort is given in Table 2. Over the random sample of 50 children over the 96 time intervals, there were 17,703 units or bubbles of data in the original file. This file had 723 false positive values or 'extra' bubbles or units of data than was provided in the 'corrected file'. Thus, the total false positive rate was 4% (723/17703*100).

Table 2: Summary of false positives based on comparing original and corrected file of 50 cases (B cohort, Wave 1)

Source	True positives	False positives	False positive rate (%)
Total	**16,980**	**723**	**4.1**
Activity	5,898	311	5.0
Where	4,235	107	2.5
Who	6,732	302	4.3
Paid	115	3	2.5

2.3 Recoding to reduce false positives

K cohort, Wave 1

A number of recodes were experimented with to reduce the rate of false positives. Table 3 gives a summary of the impact of electronic recodes on the original file in terms of both reducing the number of false positives as well as introducing false negatives into the data. These recodes reduced the false positive for the test file to 5%. The recodes are described in greater detail below.

Table 3: Summary of electronic corrections made to original file of 47 cases (4–5 year olds)

	Imputed true negatives[a]	Imputed false negatives[b]
If sleeping (recode other activities = 0)		
Early morning (4 am to 9 am)	64	6
Late evening (9 pm to 4 am)	20	6
If awake in bed (recode other activities = 0)	1	0
If one activity and implausible context data		
Travel (recode home inside or alone)	12	11
Walk/ride (recode inside a home)	11	1
Television and computer (recode outside)	1	0
If sleeping or awake in bed (recode outside)	12	0
If at day care centre outside plausible hours and give other location (recode day care)	4	0
Total activities and context data	**125**	**24**

Notes: [a] Imputed true negatives are those cells that were imputed as '0' where the corrected file had indicated that the positive response was a false positive. [b] Imputed false negatives are those cells that were imputed to '0' where the corrected file indicated that the positive response was a true positive.

Correction 1: Being asleep and doing other activities at the same time

In order to reduce the likelihood of recoding a transitional phase, the child had to be asleep in a given interval and in the preceding and following intervals. Three separate time periods were tested: morning (4 am to 9 am), nighttime (9 pm to 4 am) and daytime (9 am to 9 pm).

For the morning and nighttime periods, if an activity occurred simultaneously with sleep, non-sleep activity time was coded as zero. The recoding of activities occurring simultaneously with sleep in the morning and in the evening was relatively successful, with an aggregate number of 84 correct recodes and 12 incorrect recodes. In the full Wave 1 file (i.e. n = 7,449) this resulted in 23,216 recodes of positive responses. This correction was not performed on the Wave 2 data (n = 6,906); however, if it had been it would have only resulted in 4,498 recodes. This provides further evidence that this correction removed many more false positives than true ones.

In the period between 9 am and 9 pm, children were most likely to be periodically transitioning between sleep and other activities. Attempting to recode activities occurring simultaneously with sleep, in this period, yielded no corrections to false positives and resulted in eight true positives being recoded. Alternatively, an attempt to recode sleep resulted in recoding four false positives and five true positives. Given that both these alternatives resulted in more incorrect recodes than correct ones, neither was performed on the main data file.

Correction 2: Being awake in bed and doing other unlikely activities at the same time

Other unlikely activities are defined as:

- bathing, dressing, hair care, health care
- using computer/computer games
- walking for travel or fun
- riding bicycle, trike, etc. (travel or fun)
- other exercise – swim/dance/run about

- travelling in pusher or on bicycle seat
- travelling in car/other household vehicle
- travelling on public transport, ferry, plane
- taken places with adult (e.g. shopping)
- organised lessons activities.

If the children were doing these activities as well as being awake in bed, other activities were coded as zero. Again, in order to reduce the likelihood of recoding a transitional phase, the child had to be awake in bed in a given interval, while in the following and preceding interval they had to be either asleep or awake in bed.

The impact of recoding activities occurring simultaneously with the child being awake in bed was relatively minor, with only one false positive being recoded in the 47 diaries in the original file, and 1,225 positive responses being recoded in the full file. This is not surprising given that 4–5 year old children are not often awake in bed for long periods of time unless they are ill or are having trouble getting to sleep. In Wave 2, 1,077 positive responses would have been recoded due to this correction.

Correction 3: A child cannot be travelling and be inside at home or be alone

A child cannot be travelling (travelling in a pusher/ travelling in a car/ travelling on public transport/taken places with an adult) and be simultaneously at home inside (or in someone else's home) or be alone, if travelling was their sole activity for the time period. Recoding the context data as 0 where this occurred resulted in slightly more false positives being altered than true ones. In the full file this resulted in 4,111 positive responses being recoded. In Wave 2, 1,359 responses would have been recoded.

Correction 4: A child cannot be walking/riding and be inside a home

A child cannot be walking for travel or fun or riding a bike or trike, etc. and be inside their own or someone else's home if this is their only activity for the time period. Recoding all incidences of being inside as zero resulted in many more false positives being altered than true ones. In the full file this resulted in 1,996 positive responses being recoded. In Wave 2, only 282 responses would have been recoded.

Correction 5: A child cannot be watching television or using a computer and be outside

A child cannot be watching television or using the computer and be outside. Recoding being outside as zero in this situation resulted in only one false positive correction for these 47 diaries but no alterations to true positives. In the full file this correction resulted in the recoding of 430 positive responses. In Wave 2, 157 responses would have been recoded.

Correction 6: A child cannot be sleeping or awake in bed and be outside

If a child was awake in bed or asleep and this was their only activity they cannot be outside. While this does exclude any children who were camping (assuming a tent doesn't count as indoors), in the test cases available this correction eliminated 12 false positives without altering a true positive. In the full file this resulted in 2,459 positive responses being recoded. In Wave 2, 535 responses would have been recoded.

Correction 7: A child cannot be at a day care centre/play group outside the hours of 7 am to 7 pm

If a response was given outside of these hours it was recoded to zero. In the original file this resulted in four corrections to false positives without creating any false negatives, while in the full file 1,417 responses were recoded by this correction. In Wave 2, 610 responses would have been recoded.

B cohort, Wave 1

Table 4 gives a summary of the impact of electronic recodes on the original file for the infants in terms of both reducing the number of false positives as well as introducing false negatives into the data. In summary, the recoding resulted in a reduction of 121 of the 723 false positive data, with little creation of false negatives. As a result of these recodes, the false positive rate fell from 4% to 3%. The recodes are described in greater detail below.

Table 4: Summary of electronic corrections made to original file of 50 cases (B cohort, Wave 1)

	n	Imputed true negatives[a]	Imputed false negatives[b]
If sleeping (recode other activities = 0)			
4 am to 7 am	50	28	8
If alone/sleeping then can't be with others			
4 am to 3 pm	50	38	15
10 pm to 4 am	50	22	0
Travelling and at home or alone	50	27	9
If one activity is breastfeeding, bathing, being held or read to, recode alone	50	4	1
If at day care centre outside of the hours of 7 am to 7 pm and give other location (recode day care)	50	2	0
Total activities and context data	**50**	**121**	**33**

Notes: [a] Imputed true negatives are those cells that were imputed as '0' where the corrected file had indicated that the positive response was a false positive. [b] Imputed false negatives are those cells that were imputed to '0' where the corrected file indicated that the positive response was a true positive.

Correction 1: Being asleep and doing other activities at the same time

Children should not be asleep and also be active in another activity in the same time interval unless the child was in transition between activities. For this recode, activities that were indicated in the same time period of sleep were recoded for intervals where the child was also asleep in the preceding and following interval. This recode was tried for a number of different time periods but was only successful at the start of the day between 4 am and 7 am, where it recoded 28 false positives and eight true ones. In the full file (*n* = 7,782) this recode resulted in 11,278 positive responses being altered. As for the K cohort, these corrections were not repeated in Wave 2; however, if they had been, only 1,464 responses would have been recoded, suggesting that most Wave 1 responses recoded were false positives.

Correction 2: Being asleep alone and with someone at the same time

If the child was sleeping alone in a time period as well as the one preceding and following it, all other data in the 'in the same room as' section was recoded to zero. The only time period this didn't work for was the evening between 3 pm and 10 pm, so this period was excluded from this recode. Outside of these times it resulted in 60 corrections to false positives while introducing only 15 false negatives. In the full file this recode resulted in 29,016 responses being altered. In Wave 2, 8,618 responses would have been recoded.

Correction 3: A child cannot be travelling and be inside at home or be alone

A child cannot be travelling (travelling in a pusher/ travelling in a car/ travelling on public transport/taken places with an adult) and be simultaneously at home inside or be alone. A child was identified as travelling if they are travelling in a given interval and the preceding and following interval. In this situation, the 'at home' or 'alone' response would be removed. This correction removed 27 false positives while introducing nine false negatives. In the full file, it led to 6,303 positive responses being removed. In Wave 2, 1,098 positive responses would have been removed.

Correction 4: Being alone and with improbable activities

If a child's only activities are breastfeeding, being held, having personal grooming tasks performed, or being read a story or talked or sung to, any response that the child was alone for the period was recoded to zero. Correcting this removed four false positives and produced one false negative, while in the full file this recode led to the removal of 1,031 responses of 'alone'. In Wave 2, 420 responses would have been removed.

Correction 5: Being at a day care centre/playgroup at improbable hours

Any response indicating that the child was at a day care centre outside the hours of 7 am to 7 pm was recoded to zero. This recode corrected two false positives and produced no false negatives, while in the full file, 593 positive responses were removed. In the Wave 2 file, 301 responses would have been removed.

2.4 Coding to improve data quality

A number of further recodes were undertaken to improve other aspects of data quality, such as reducing missing or contradictory data.

B and K cohorts

These operations were performed on all three waves of diary data for both cohorts.

Improvement 1: Recoding not sure when other activities given in the same time interval

Ideally, respondents should only have given 'not sure' as a response if they were unable to report any of the child's activities in a 15-minute block. Where this has happened the 'not sure' response was coded to zero. For Wave 1, this removed 12,770 'not sure' responses from the K cohort file and 10,026 'not sure' responses from the full infant file. For Wave 2, these figures were 4,678 for the K cohort and 3,717 for the B cohort and, in Wave 3, they were 5,380 for the K cohort and 3,759 for the B cohort.

Improvement 2: Imputing not sure or missing activity data as sleep and the early morning

If the parent was not sure of what the child was doing or activity data was missing in the early morning (4 am to 9 am) and the sequence of not sure/missing ended with either the child being awake in bed or sleeping, the not sure/missing was recoded as sleep. In Wave 1, these changes created an extra 984 sleep responses in the full K cohort file and 1,292 extra sleep responses in the full B cohort file. In Wave 2, these figures were 1,036 for the K cohort and 1,002 for the B cohort and, in Wave 3, they were 943 for the K cohort and 566 for the B cohort.

Improvement 3: Imputing not sure or missing activity data as sleep at nighttime

If the parent was not sure of what the child was doing or activity data was missing at nighttime (9 pm to 4 am) and this sequence began following the child being either awake in bed or sleeping, the not sure/missing data was recoded as sleep. This created an extra 2,540 sleep responses in the full K cohort file and 4,101 in the full B cohort file. In Wave 2, the figures were 2,681 for the K cohort and 3,933 for the B cohort and, in Wave 3, they were 2,229 for the K cohort and 2,328 for the B cohort.

Improvement 4: Other missing data

If there was a single time period with missing activity data and the child remained in the same location, then either the activity before or after the missing bubble was randomly allocated to the missing bubble. This improvement imputed activities in 2,517 time periods in the full K cohort file and 3,022 time periods in the B cohort file. For Wave 2, these figures were 1,956 for the K cohort and 2,389 for the B cohort and, in Wave 3, they were 1,553 for the K cohort and 1,767 for the B cohort.

Improvement 5: Missing 'who' information in child care

If a child's 'where' information includes 'day care centre/playgroup', it can reasonably be assumed they are in the presence of other children and other adults when alternative information is missing. This improvement imputed data in 17,294 time periods in the full K cohort file and 3,169 time periods in the B cohort file. As might be expected given the rise in time in non-parental care for the children as they get older, these numbers were higher in Wave 2 with 35,097 time periods for the K cohort and 12,712 for the B cohort and, in Wave 3, they were 23,045 for the K cohort and 15,835 for the B cohort.

2.5 Exclusion of cases

It is common practice when analysing time use diary data to exclude cases with poor quality data, often indicated by rules of thumb such as more than 90 minutes of missing information (e.g. Egerton and Gershuny, 2004; Fisher, 2002). The LSAC time use diaries use a different response format than many other similar instruments (i.e. the use of scanned responses rather than coding of text responses) and this may have an effect on the quality of the diary data and on which cases should be excluded. Cases considered to be of poor quality were removed from the main diary dataset and placed in a separate file so that they could be re-included for any analysis where the user thought they might be valuable.

2.6 Criteria for exclusion

Three criteria were used to exclude cases from the dataset.

Cases with large amounts of missing data

As mentioned above, it is common practice to remove cases with more than 90 minutes missing activity data from analyses. However, analyses of the LSAC data suggested that using this rule of thumb might be inappropriate as children who spent time away from their parents (e.g. in child care) were more likely to have greater levels of missing activity data. Instead, a diary was deleted from the file if it had no data of any kind for more than 90 minutes (or six time intervals). In Wave 1, this criterion excluded 239 diaries (3%) from the B cohort file and 368 diaries (5%) from the K cohort file. For Wave 2, 235 (4%) diaries were deleted from the B cohort file and 268 (4%) were deleted from the K cohort file. For Wave 3, 147 (3%) diaries were deleted from the B cohort file and 233 (4%) were deleted from the K cohort file.

Cases with large numbers of simultaneous activities

Most time use diaries request respondents to describe their main activity for each time period, with limited opportunities to describe secondary activities. The format of the LSAC time use diary meant that a number of activities could be specified separately; however, where numbers were large, it often indicated that the respondent had trouble understanding the task. As such, it was decided to exclude any respondent that gave more than five simultaneous activities for more than six time periods. In Wave 1, this criterion excluded 78 diaries (1%) from the B cohort file and 55 diaries (1%) from the K cohort file, while in Wave 2, 26 (0.4%) B cohort diaries and 16 (0.2%) K cohort diaries were deleted. For Wave 3, 11 (0.2%) diaries were deleted from the B cohort file and 11 (0.2%) were deleted from the K cohort file.

Cases with few changes in activities

Diaries with few changes in activities tended to occur when the parent either did not have a good idea of the child's activity (e.g. large amount of time in non-parental care) or was not able to fill in the diary in detail. It was decided that fewer than 10 different activities over the 24-hour period represented an unacceptable lack of detail. This excluded 59 diaries (1%) from the B cohort file and 144 diaries (2%) from K cohort file. In Wave 2, these figures were 110 (2%) and 171 (3%) respectively. For Wave 3, 120 (2%) diaries were deleted from the B cohort file and 159 (3%) were deleted from the K cohort file.

There were some diaries excluded for more than one reason, so in total for Wave 1, 330 diaries (4%) were excluded from the B cohort file and 490 (7%) were excluded from the K cohort file. The effect of the exclusion of these diaries on the socio-demographic composition of the time use diary sample can be seen in Table 5 (on page 10). The deleted diaries tended to come from lower socio-economic status families.

Table 5: Effect of deleting problem cases on socio-demographic composition of the Wave 1 LSAC TUD sample (unweighted data)

Wave 1	B cohort			K cohort		
	Full LSAC sample (%)	Full TUD sample (%)	Reduced TUD sample (%)	Full LSAC sample (%)	Full TUD sample (%)	Reduced TUD sample (%)
Gender						
Male	51.2	51.4	51.6	50.9	51.2	51.6
Female	48.8	48.6	48.4	49.1	48.8	48.4
Age range of children (B cohort/K cohort)						
3–5 months/ 51–53 months	11.2	11.1	11.2	10.6	10.6	10.7
6–11 months/ 54–59 months	73.2	73.3	73.4	72.1	72.1	72.8
12–14 months/ 60–62 months	14.7	14.8	14.8	16.1	16.1	15.7
15–19 months/ 63–67 months	1.0	0.8	0.6	1.3	1.3	0.8
Family type						
Couple family:	90.7	91.5	93.0	86.0	87.0	88.9
both biological	90.1	91.0	92.5	82.9	84.3	86.6
other (e.g. step/blended)	0.6	0.6	0.5	3.1	2.8	2.3
Single parent family	9.3	8.5	7.0	14.0	12.9	11.0
Siblings						
Only child	39.5	40.0	40.7	11.5	11.2	10.8
One sibling	36.8	36.8	36.9	48.4	49.4	50.7
Two or more siblings	23.7	23.2	22.3	40.1	39.4	38.9
Cultural background						
Aboriginal or Torres Strait Islander	4.5	3.9	2.7	3.8	3.2	2.4
Mother speaks a language other than English at home	14.5	13.4	11.2	15.7	14.7	12.3
Work status						
Both parents or lone parent work/s	47.9	48.8	50.5	55.5	56.0	57.2
One parent works (in couple family)	40.8	41.2	41.5	32.8	33.5	34.2
No parent works	11.3	10.1	8.0	11.6	10.5	8.6
Educational status						
Mother completed Year 12	66.9	68.7	71.9	58.6	60.2	63.0
Father completed Year 12	58.5	59.4	60.8	52.7	53.3	54.6
Child care						
Child has a regular care arrangement (including school)	35.9	35.7	35.4	96.7	97.1	97.6
State						
New South Wales	31.6	30.8	30.0	31.6	31.2	30.8
Victoria	24.5	24.5	24.5	25.0	25.0	24.9
Queensland	20.6	20.9	21.2	19.8	20.1	20.4
South Australia	6.8	6.6	6.4	6.8	6.6	6.2
Western Australia	10.4	10.8	11.2	10.2	10.4	10.8
Tasmania	2.2	2.3	2.4	2.7	2.8	2.9
Northern Territory	1.7	1.8	1.8	1.7	1.6	1.5
Australian Capital Territory	2.1	2.3	2.4	2.3	2.3	2.4

Table continued on next page →

Wave 1	B cohort			K cohort		
	Full LSAC sample (%)	Full TUD sample (%)	Reduced TUD sample (%)	Full LSAC sample (%)	Full TUD sample (%)	Reduced TUD sample (%)
Region						
Capital city statistical division	62.5	62.8	63.1	62.1	62.0	61.6
Balance of state	37.5	37.2	36.9	37.9	38.0	38.4
Number of observations (*n*) [a]	**5,107**	**8,858**	**7,452**	**4,983**	**8,565**	**6,959**

Note: [a] TUD samples are larger than the LSAC sample as respondents were asked to complete two diaries.

In Wave 2, 335 (5%) were deleted from the B cohort file and 405 (6%) from the K cohort file. The effect of the exclusion of these diaries on the socio-demographic composition of the time use diary sample can be seen in Table 6. Again, the deleted diaries tended to come from lower socio-economic status families.

Table 6: Effect of deleting problem cases on socio-demographic composition of the Wave 2 LSAC TUD sample (unweighted data)

Wave 2	B cohort			K cohort		
	Full LSAC sample (%)	Full TUD sample (%)	Reduced TUD sample (%)	Full LSAC sample (%)	Full TUD sample (%)	Reduced TUD sample (%)
Gender						
Male	51.1	51.3	51.2	51.0	51.7	51.9
Female	48.9	48.7	48.8	49.0	48.3	48.2
Age range of children (B cohort/K cohort)						
27–32 months/ 75–77 months	6.3	6.6	6.7	7.1	7.5	7.6
30–35 months/ 78–83 months	64.8	66.1	66.5	63.7	64.7	64.9
36–38 months/ 84–86 months	23.5	23.0	22.8	23.8	23.3	23.2
39–43 months/ 87–91 months	5.4	4.3	3.9	5.4	4.5	4.3
Family type						
Couple family:	89.0	91.6	92.0	85.2	88.2	88.9
both biological	88.0	90.5	91.0	81.3	85.2	85.9
other (e.g. step/blended)	1.0	1.1	1.1	3.9	3.1	2.9
Single parent family	11.0	8.4	8.0	14.8	11.8	11.1
Siblings						
Only child	19.3	19.1	18.9	9.1	8.7	8.8
One sibling	49.1	51.4	51.9	45.2	47.7	48.1
Two or more siblings	31.6	29.6	29.2	45.7	43.6	43.1
Cultural background						
Aboriginal or Torres Strait Islander	3.9	2.5	2.3	3.4	2.3	2.3
Mother speaks a language other than English at home	13.4	11.8	11.1	14.7	13.6	12.5
Work status						
Both parents or lone parent work/s	56.9	58.0	58.5	65.4	67.6	68.3
One parent works (in couple family)	33.8	35.0	35.0	26.1	26.0	25.8
No parent works	9.3	7.0	6.5	8.6	6.5	5.9
Educational status						
Mother completed Year 12	69.0	73.0	74.1	60.1	63.8	64.9
Father completed Year 12	59.7	62.2	62.6	53.2	55.5	56.0

Table continued on next page →

Wave 2	B cohort			K cohort		
	Full LSAC sample (%)	Full TUD sample (%)	Reduced TUD sample (%)	Full LSAC sample (%)	Full TUD sample (%)	Reduced TUD sample (%)
Child care						
Child has a regular care arrangement (including school)	70.4	71.3	71.5	99.7	99.7	99.6
State						
New South Wales	31.1	30.9	30.9	31.1	30.9	31.0
Victoria	24.3	24.8	24.7	24.3	24.5	24.2
Queensland	21.4	21.1	21.1	21.4	20.4	20.8
South Australia	6.7	6.8	6.8	6.7	6.9	6.8
Western Australia	10.6	10.4	10.5	10.6	10.4	10.4
Tasmania	2.3	2.3	2.4	2.3	3.2	3.2
Northern Territory	1.4	1.3	1.3	1.4	1.3	1.2
Australian Capital Territory	2.2	2.3	2.3	2.2	2.4	2.4
Region						
Capital city statistical division	61.9	62.6	62.5	61.6	61.4	61.4
Balance of state	38.1	37.4	37.5	38.4	38.6	38.6
Number of observations (*n*) [a]	**4,606**	**6,917**	**6,582**	**4,464**	**6,858**	**6,483**

Note: [a] TUD samples are larger than the LSAC sample as respondents were asked to complete two diaries.

In Wave 3, 228 (4%) were deleted from the B cohort file and 339 (6%) from the K cohort file. The effect of the exclusion of these diaries on the socio-demographic composition of the time use diary sample can be seen in Table 7. Again, the deleted diaries tended to come from lower socio-economic status families.

Table 7: Effect of deleting problem cases on socio-demographic composition of the Wave 3 LSAC TUD sample (unweighted data)

Wave 3	B cohort			K cohort		
	Full LSAC sample (%)	Full TUD sample (%)	Reduced TUD sample (%)	Full LSAC sample (%)	Full TUD sample (%)	Reduced TUD sample (%)
Gender						
Male	51.3	52.1	52.1	51.1	51.4	51.2
Female	48.7	47.9	47.9	48.9	48.6	48.8
Age range of children (B cohort/K cohort)						
27–32 months/ 75–77 months	7.8	8.4	8.5	8.4	8.8	8.8
30–35 months/ 78–83 months	67.2	68.5	68.5	65.7	65.1	65.1
36–38 months/ 84–86 months	20.7	19.9	19.8	21.9	22.5	22.5
39–43 months/ 87–91 months	4.3	3.1	3.1	4.1	3.7	3.6
Family type						
Couple family:	88.9	91.6	91.8	85.6	86.4	86.7
both biological	85.8	89.0	89.3	78.8	79.7	80.2
other (e.g. step/blended)	3.0	2.6	2.3	6.8	6.7	6.5
Single parent family	11.1	8.4	8.2	14.4	13.6	13.3
Siblings						
Only child	10.4	10.4	10.2	8.2	8.6	8.6
One sibling	48.1	51.2	51.6	44.1	45.2	45.2
Two or more siblings	41.5	38.4	38.2	47.7	46.2	46.2

Table continued on next page →

Wave 3	B cohort			K cohort		
	Full LSAC sample (%)	Full TUD sample (%)	Reduced TUD sample (%)	Full LSAC sample (%)	Full TUD sample (%)	Reduced TUD sample (%)
Cultural background						
Aboriginal or Torres Strait Islander	3.4	2.0	2.0	2.9	2.6	2.4
Mother speaks a language other than English at home	12.6	10.7	10.2	13.8	13.5	13.2
Work status						
Both parents or lone parent work/s	63.0	64.0	64.2	72.8	73.1	73.4
One parent works (in couple family)	29.7	30.9	30.9	20.7	20.7	20.7
No parent works	7.4	5.1	4.9	6.5	6.2	5.9
Educational status						
Mother completed Year 12	69.8	74.1	74.8	61.3	62.2	62.8
Father completed Year 12	60.4	63.3	63.7	54.0	54.9	55.1
Child care						
Child has a regular care arrangement (including school)	96.6	97.4	97.4	99.5	99.4	99.4
State						
New South Wales	30.1	29.6	29.6	30.8	31.2	31.5
Victoria	24.6	24.3	24.0	24.4	19.6	19.5
Queensland	22.0	21.4	21.5	20.8	23.2	23.1
South Australia	7.0	7.6	7.7	6.9	5.1	5.1
Western Australia	10.3	10.7	10.7	10.2	14.9	14.9
Tasmania	2.4	2.6	2.6	30.	2.4	2.4
Northern Territory	1.2	1.3	1.3	1.4	1.6	1.5
Australian Capital Territory	2.4	2.6	2.6	2.5	2.0	2.1
Region						
Capital city statistical division	61.9	62.6	62.6	61.3	61.3	61.0
Balance of state	38.2	37.4	37.4	38.7	38.8	39.0
Number of observations (*n*)[a]	**4,384**	**5,909**	**6,582**	**4,332**	**5,924**	**5,585**

Note: [a] TUD samples are larger than the LSAC sample as respondents were asked to complete two diaries.

2.7 Summary

Corrections to improve data quality and deletion of problem cases had some effect on the rate of false positives due to scanning errors in the corrected files (i.e. the cases that had been checked against the paper forms). When these improvements were performed on the corrected file, the false positive rate dropped to 5% for the K cohort file and 3% for the B cohort file. In Wave 2, these same recodes, when applied to more rigorously checked, scanned files, recoded far fewer responses. This adds further evidence that it was largely false positive responses that were being recoded in Wave 1. Tables 8a to 8f show the effect of the recodes on estimates produced by the full data file. In the final Wave 1 B cohort file (i.e. with cases deleted and all corrections made) 88% of cases had at least one correction made to them, while in the final K cohort file, 84% of cases had at least one correction. In Wave 2, these proportions were much lower: 41% for the B cohort and 48% for the K cohort. In Wave 3, they were 42% for the B cohort and 46% for the K cohort.

Table 8a: Effect of recoding and case deletions on estimates of time use in number of minutes/day (B cohort, Wave 1)[a]

	Raw file	After recodes	After recodes and deletions
What was the child doing?			
Not sure what child was doing	42.5	19.8	19.0
Sleeping, napping	772.5	784.5	800.3
Awake in bed/cot	48.2	46.9	45.2
Looking around, doing nothing	28.2	27.6	25.8
Bathing/nappy change, dress/hair care	91.5	91.3	90.9
Breastfeeding	53.7	52.9	52.3
Other eating, drinking, being fed	129.4	128.9	129.1
Crying, upset	43.4	42.9	42.1
Destroying things, creating mess	22.6	22.3	21.1
Held, cuddled, comforted, soothed	129.8	127.5	128.0
Watching TV, video or DVD	37.3	37.3	36.0
Listening to tapes, CDs, radio, music	29.3	28.4	26.9
Read a story, talked/sung to, sing/talk	77.9	77.1	76.2
Colour/drawing, look at book, puzzles	8.5	8.5	7.8
Organised activities/playgroup	10.6	10.5	9.6
Crawl, climb, swing arms or legs	121.7	120.2	119.2
Other play, other activities	137.4	135.7	137.9
Visiting people, special event, party	41.2	40.3	39.1
Taken places with adult (e.g. shopping)	54.7	54.1	53.2
Travel			
Taken out in pram or bicycle seat	36.8	36.4	35.3
Travel in car/other household vehicle	54.9	54.4	54.2
Travel on public transport, ferry, plane	3.0	2.9	2.1
Where was the child?			
Own home (indoors)	1111.0	1100.9	1130.0
Other person's home (indoors)	68.4	68.4	68.7
Day care centre/playgroup	22.2	20.9	19.7
Other indoors	43.7	43.7	43.6
Other outdoors	66.3	66.3	67.2
In the same room, nearby if outside			
Alone	370.5	366.4	380.2
Mother, step-mother	781.6	756.3	775.7
Father, step-father	405.2	393.9	403.7
Grandparent(s)/other adult relative	105.0	103.0	102.9
Brother(s), sister(s), other children	365.5	364.4	371.5
Other adult(s)	65.2	71.8	71.7
Dog, cat or other pet (not fish)	119.4	114.3	116.3
Payment			
Someone paid for this activity	21.9	21.9	21.4

Note: [a] Analysis uses weights that adjust for general LSAC non-response as well as weighting each day of the week equally in the analysis. These weights are recalculated when the poor-quality cases are deleted.

Table 8b: Effect of recoding and case deletions on estimates of time use in number of minutes/day (K cohort, Wave 1) [a]

	Raw file	After recodes	After recodes and deletions
What was the child doing?			
Not sure what child was doing	68.4	39.0	38.4
Sleeping, napping	626.6	634.7	650.8
Awake in bed	37.3	33.8	34.0
Eating, drinking, being fed	124.0	120.7	123.0
Bathing, dressing, hair care, health care	59.9	58.8	59.6
Do nothing, bored/restless	7.7	7.0	6.5
Crying, upset, tantrum	10.1	9.1	8.9
Destroy things, create mess	8.0	6.9	6.3
Held, cuddled, comforted, soothed	40.5	37.5	37.0
Being reprimanded, corrected	12.3	12.0	11.4
Watching TV, video, DVD, movie	124.8	121.1	122.6
Listening to tapes, CDs, radio, music	18.9	17.9	17.5
Use computer/computer games	16.6	15.9	15.5
Read a story, talk/sing, talked/sung to	61.3	59.7	59.8
Colour, look at book, educational game	43.0	42.2	41.7
Being taught to do chores, read, etc.	18.3	17.5	17.1
Walk for travel or for fun	13.7	13.1	12.8
Ride bicycle, trike, etc. (travel or fun)	17.9	17.1	16.5
Other exercise – swim /dance/run about	47.4	45.7	46.4
Visiting people, special event, party	50.4	46.9	47.2
Other play, other activities	109.6	105.1	108.1
Travel in pusher or on bicycle seat	3.9	3.4	3.0
Travel in car/other household vehicle	57.6	55.4	56.6
Travel on public transport, ferry, plane	5.5	4.7	4.3
Taken places with adult (e.g. shopping)	49.5	47.4	47.9
Organised lessons/activities	73.8	71.9	74.5
Where was the child?			
Own home (indoors)	917.4	906.5	941.2
Other person's home (indoors)	65.4	64.9	64.7
Day care centre/playgroup	108.6	104.9	106.3
Other indoors	73.0	73.0	75.6
Other outdoors	107.3	102.2	105.2
In the same room, nearby if outside			
Alone	226.4	225.8	236.9
Mother, step-mother	593.4	593.4	616.3
Father, step-father	343.5	343.5	357.7
Grandparent(s)/other adult relative	92.8	92.8	92.7
Brother(s), sister(s), other children	662.8	711.0	738.8
Other adult(s)	123.8	172.0	175.5
Dog, cat or other pet (not fish)	127.8	127.8	131.7
Payment			
Someone paid for this activity	74.1	74.1	77.2

Note: [a] Analysis uses weights that adjust for general LSAC non-response as well as weighting each day of the week equally in the analysis. These weights are recalculated when the poor-quality cases are deleted.

Table 8c: Effect of recoding and cases deletions on estimates of time use in number of minutes/day (B cohort, Wave 2)[a]

	Raw file	After recodes	After recodes and deletions
What was the child doing?			
Not sure what child was doing	60.2	48.6	47.4
Sleeping, napping	662.3	673.4	686.6
Awake in bed	40.7	41.0	41.0
Eating, drinking, being fed	117.7	119.1	120.6
Bathing, dressing, hair care, health care	54.0	54.7	55.2
Doing nothing, bored/restless	5.9	6.0	5.7
Crying, upset, tantrum	12.1	12.3	12.3
Arguing, fighting	5.6	5.7	5.4
Destroy things, create mess	7.1	7.2	6.5
Being reprimanded	9.7	9.8	9.1
Being held, cuddled, comforted, soothed	45.8	46.2	45.8
Watching TV, video, DVD, movie	94.5	95.1	95.9
Listening to tapes, CDs, radio, music	20.4	20.5	20.6
Using computer, computer game	4.3	4.3	4.1
Read a story, told a story, sung to	33.6	33.9	34.3
Colour/draw, look at book, educational game	36.2	36.5	36.3
Quiet free play	76.6	76.9	78.7
Active free play	88.8	89.1	90.4
Being taught to do chores	11.7	11.7	11.4
Visiting people, special event, party	72.0	72.1	73.4
Organised lessons/activities	15.8	15.8	16.0
Travel			
Walking	13.0	13.1	12.5
Ride bicycle/trike	9.4	9.5	9.0
Travel in car	51.7	52.0	52.0
Travel in a pusher/bicycle seat	5.3	5.3	5.1
Travel on public transport	1.5	1.5	1.3
Taken places with adult (e.g. shopping)	34.5	34.6	34.7
Where was the child?			
Own home (indoors)	944.8	944.8	974.8
Other person's home (indoors)	61.2	61.2	61.7
Day care centre/playgroup	87.5	87.5	85.4
Other indoors	85.5	85.5	86.9
Other outdoors	68.6	68.6	70.4
In the same room, nearby if outside			
Alone	298.3	298.3	311.2
Mother, step-mother	677.7	677.7	699.3
Father, step-father	368.2	368.2	379.2
Grandparent(s)/other adult relative	94.0	94.0	94.9
Brother(s), sister(s), other children	551.3	590.0	606.4

Table continued on next page →

	Raw file	After recodes	After recodes and deletions
Other adult(s)	91.1	129.8	129.0
Dog, cat or other pet (not fish)	113.9	113.9	118.3
Payment			
Someone paid for this activity	53.3	53.3	54.3

Note: [a] Analysis uses weights that adjust for general LSAC non-response as well as weighting each day of the week equally in the analysis. These weights are recalculated when the poor-quality cases are deleted.

Table 8d: Effect of recoding and case deletions on estimates of time use in number of minutes/day (K cohort, Wave 2)[a]

	Raw file	After recodes	After recodes and deletions
What was the child doing?			
Not sure what child was doing	99.2	86.6	86.1
Sleeping, napping	598.4	607.5	620.2
Awake in bed	30.8	31.1	30.5
Eating and drinking	95.8	96.9	98.7
Bathing, dressing, hair care, health care	49.6	50.2	51.0
Do nothing, bored/restless	4.1	4.1	3.7
Crying, upset, tantrum	3.0	3.0	2.8
Arguing, fighting, destroy things	3.8	3.8	3.5
Held, cuddled, comforted, soothed	18.9	19.1	18.3
Being reprimanded, corrected	6.8	6.9	6.6
Watching TV, video, DVD, movie	91.2	91.7	92.9
Listening to tapes, CDs, radio, music	12.3	12.4	12.2
Use computer/computer games	18.4	18.6	18.5
Read a story, talk/sing, talked/sung to	16.9	17.1	17.1
Reading looking at book by self	21.0	21.2	21.3
Quiet free play	47.8	47.9	49.5
Active free play	65.6	65.8	67.6
Helping with chores/jobs	19.0	19.2	19.0
Visiting people, special event, party	59.7	59.8	59.9
Organised sport/physical activity	18.7	18.8	18.8
Other organised lessons/activities	17.3	17.4	18.0
Travel			
Walk for travel or for fun	9.6	9.6	9.4
Ride bicycle, trike, etc. (travel or fun)	11.4	11.5	11.5
Travel in car	47.8	48.0	49.0
Travel on public transport	4.7	4.7	4.7
Taken places with adult (e.g. shopping)	20.4	20.5	20.7
Where was the child?			
Own home (indoors)	784.2	784.2	812.8
Own home (outdoors)	56.7	56.7	57.0
School, after/before school care	219.8	219.8	223.1
Other indoors	78.1	78.1	78.1
Other outdoors	65.5	65.5	67.3

Table continued on next page →

	Raw file	After recodes	After recodes and deletions
In the same room, nearby if outside			
Alone	247.4	247.4	259.3
Mother, step-mother	479.1	479.1	494.1
Father, step-father	303.3	303.3	312.7
Grandparent(s)/other adult relative	65.5	65.5	65.4
Brother(s), sister(s), other children	649.3	759.6	781.9
Other adult(s)	141.6	251.9	257.2
Dog, cat or other pet (not fish)	126.2	126.2	130.2
Homework			
Activity done as part of homework	11.7	11.7	11.5

Note: [a] Analysis uses weights that adjust for general LSAC non-response as well as weighting each day of the week equally in the analysis. These weights are recalculated when the poor-quality cases are deleted.

Table 8e: Effect of recoding and cases deletions on estimates of time use in number of minutes/day (B cohort, Wave 3)[a]

	Raw file	After recodes	After recodes and deletions
What was the child doing?			
Not sure what child was doing	60.0	48.5	47.7
Sleeping, napping	626.8	635.2	646.9
Awake in bed	35.9	36.2	35.9
Eating, drinking, being fed	106.5	107.7	109.5
Bathing, dressing, hair care, health care	53.4	54.0	55.0
Doing nothing, bored/restless	3.7	3.8	3.6
Crying, upset, tantrum	4.8	4.9	4.9
Arguing, fighting	5.7	5.8	5.7
Destroy things, create mess	3.3	3.3	3.2
Being reprimanded, corrected	6.6	6.7	6.6
Being held, cuddled, comforted, soothed	28.4	28.7	29.3
Watching TV, video, DVD, movie	99.1	99.7	100.9
Listening to tapes, CDs, radio, music	15.9	16.0	15.9
Using computer, computer game	14.0	14.1	14.3
Read a story, told a story, sung to	45.9	46.3	47.2
Colour/draw, look at book, educational game	39.4	39.6	39.9
Quiet free play	65.5	65.8	67.3
Active free play	82.0	82.3	83.8
Being taught to do chores	15.3	15.4	15.4
Visiting people, special event, party	63.5	63.5	64.2
Organised lessons/activities	54.5	54.5	55.6
Travel			
Walking	9.7	9.8	9.7
Travel in a pusher/bicycle seat	2.4	2.4	2.3
Travel in car	49.5	49.8	50.7
Travel on public transport	3.3	3.3	3.3

Table continued on next page →

	Raw file	After recodes	After recodes and deletions
Taken places with adult (e.g. shopping)	24.8	24.8	25.1
Ride bicycle/trike, etc.	8.1	8.1	8.1
Where was the child?			
Own home (indoors)	906.1	906.1	929.9
Other person's home (indoors)	57.8	57.8	57.3
Day care centre/playgroup/pre-school/school	151.9	151.9	152.7
Other indoors	43.4	43.4	44.4
Other outdoors	86.8	86.8	89.2
In the same room, nearby if outside			
Alone	212.7	212.7	219.0
Mother, step-mother	690.5	690.5	707.8
Father, step-father	412.6	412.6	422.6
Grandparent(s)/other adult relative	85.4	85.4	87.1
Brother(s), sister(s), other children	719.0	778.1	798.9
Other adult(s)	132.6	191.7	194.0
Dog, cat or other pet (not fish)	160.3	160.3	165.0
Payment			
Someone paid for this activity	67.3	67.3	67.8

Note: [a] Analysis uses weights that adjust for general LSAC non-response as well as weighting each day of the week equally in the analysis. These weights are recalculated when the poor-quality cases are deleted.

Table 8f: Effect of recoding and case deletions on estimates of time use in number of minutes/day (K cohort, Wave 3)[a]

	Raw file	After recodes	After recodes and deletions
What was the child doing?			
Not sure what child was doing	91.2	75.4	75.1
Sleeping, napping	584.1	592.4	606.6
Awake in bed	34.6	34.8	34.2
Eating and drinking	95.8	96.7	97.8
Bathing, dressing, hair care, health care	48.8	49.4	50.0
Do nothing, bored/restless	5.1	5.2	4.3
Sulking, upset	3.0	3.0	2.5
Arguing, fighting	5.2	5.3	4.8
Being hugged, comforted, etc.	11.6	11.7	11.2
Being reprimanded, corrected	6.8	6.9	6.2
Watching TV, video, DVD, movie	103.5	104.1	105.0
Listening to tapes, CDs, radio, music	11.9	12.0	11.3
Use computer/computer games	30.2	30.4	30.2
Read a story, talk/sing, talked/sung to	9.9	10.0	9.4
Reading /looking at book by self	24.7	24.9	25.0
Quiet free play	42.6	42.8	43.7
Active free play	62.6	62.8	64.0
Helping with chores/jobs	22.8	22.9	22.8

Table continued on next page →

	Raw file	After recodes	After recodes and deletions
Visiting people, special event, party	60.6	60.7	60.9
Organised sport/physical activity	25.4	25.4	25.9
Other organised lessons/activities	22.5	22.6	23.0
Travel			
Walk for travel or for fun	10.1	10.2	9.8
Ride bicycle, trike, etc. (travel or fun)	12.0	12.0	11.3
Travel in car	49.0	49.3	49.2
Travel on public transport	5.5	5.5	5.5
Taken places with adult (e.g. shopping)	20.9	21.0	20.6
Where was the child?			
Own home (indoors)	809.9	809.9	838.1
Own home (outdoors)	53.6	53.6	53.4
School, after/before school care	209.2	209.2	215.5
Other indoors	91.3	91.3	92.6
Other outdoors	65.7	65.7	67.4
In the same room, nearby if outside			
Alone	247.3	247.3	260.6
Mother, step-mother	524.3	524.3	542.9
Father, step-father	344.4	344.4	355.8
Grandparent(s)/other adult relative	65.1	65.1	66.0
Brother(s), sister(s)	599.6	599.6	622.0
Other children	219.1	306.1	316.6
Other adult(s)	161.8	248.8	256.7
Dog, cat or other pet (not fish)	165.5	165.5	172.6
Homework			
Activity done as part of homework	14.1	14.1	13.7

Note: [a] Analysis uses weights that adjust for general LSAC non-response as well as weighting each day of the week equally in the analysis. These weights are recalculated when the poor-quality cases are deleted.

Acknowledgement

This chapter is largely based on the work of Jude Brown and Michael Bittman of the University of New England. David Zago of the Australian Bureau of Statistics provided the information on the process used to scan Waves 2 and 3 forms.

3 Report on Adapted PPVT-III and 'Who Am I?'

3.1 Wave 1 scoring

The first Wave of the Longitudinal Study of Australian Children (LSAC) used two tests with the four-year-old sample. The Adapted PPVT-III is a shortened version of the *Peabody Picture Vocabulary Test*, Third Edition (Dunn & Dunn, 1997), which is a test of receptive vocabulary used as a screening test of verbal ability. This adaptation is based on work done in the USA for the Head Start Impact Study, with a number of changes for use in Australia. 'Who Am I?' (de Lemos & Doig, 2000) assesses the cognitive processes that underlie the learning of early literacy and numeracy skills. One item was added to the standard 'Who Am I?' booklet for use in LSAC. Summary statistics for each test are shown in Table 9.

Table 9: Summary statistics for administration of the adapted PPVT-III and 'Who Am I?' tests as part of LSAC Wave 1

	Adapted PPVT	'Who Am I?'
Number of cases	4407	4827
Mean scaled scores	64.2 (se = 0.123)	63.8 (se = 0.125)
Mean number of items correct/mean raw score	28.2 (se = 0.086)	25.6 (se = 0.103)
Minimum number of items correct	2	0
Maximum number of items correct	40	44
Reliability	0.76	0.89

Note: For the adapted PPVT-III, it was assumed that children who were not required to answer 10 'basal' items had answered these items correctly. Reliability reported here is the person separation reliability (Wright & Masters, 1982).

Adapted PPVT-III

The PPVT-III was adapted for use in LSAC by altering the administration procedures, reducing the number of items administered during testing. To determine which items to retain for the adapted version, 215 children aged from 41 to 66 months (mean = 54.7 months) were given the PPVT-III, with test administrators following standard procedures. After testing, a one-parameter (Rasch) item response model was fitted to the data, which consisted of correct and incorrect responses. The person separation reliability was 0.88. After determining the 'best' 40 items for use in a shortened version, the remaining items were then fit again to a one-parameter item response model; the person separation reliability decreased to 0.78.

Development of the model suggested that 37% of children would require only the core set of items, 5% would require the core and basal sets, and 58% would require the core and ceiling sets, resulting in an average of 26.3 items administered. The Pearson product-moment correlation between the full PPVT-III and the adapted PPVT-III was 0.93 for all children, and 0.91 for four year olds (Rothman, 2013).

Scaling

The adapted PPVT-III was scaled using a two-stage process. In the first stage, only the core set of 20 items was used, as these items had been administered to all children. For these core items, Rasch estimates were determined for each item, providing an indication of their difficulty. In the second stage, all 40 items were fitted, using the item estimates for the core items as anchors. This gave item estimates for the basal and ceiling items

relative to the core items. The final case estimates were then transformed to a scale with a mean of 64 and standard deviation of 8.

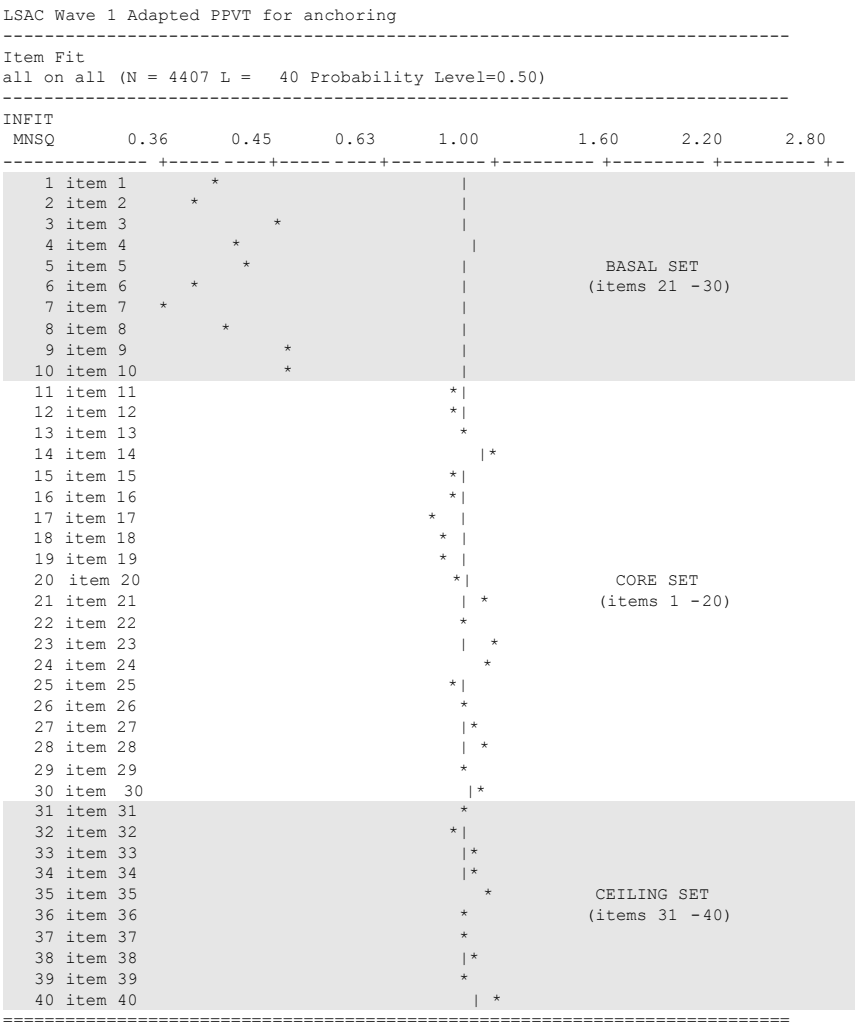
Results

In Wave 1, 4,407 children were administered the adapted PPVT-III. Children ranged in age from 43 months to 79 months (mean = 57.3); 21% were aged 60 months or older. Twenty-one per cent of children were administered only the core set, 1% were given the core and basal sets, and 78% were given the core and ceiling sets, resulting in an average of 27.9 items. The test had a person separation reliability of 0.76.

Quality of the PPVT test

The statistics indicate that the core and ceiling items used for the adapted PPVT-III test fit the Rasch model well. This is shown in Figure 3, the item fit map. The infit mean square ranged from 0.86 to 1.17 for items 11–30 (the core set) and items 31–40 (the ceiling set). On each of the items in the basal set (items 1–10), the infit mean square was extremely low (0.49 or less) because only 30 children (1%) were administered these items; all other children were assumed to have correctly answered these items. The item map in Figure 3, which shows the item estimates (difficulties) mapped against the case estimates (children's ability levels), shows that the basal items were appropriate for children given that set but that the core and ceiling items were relatively easy for those who were given those sets.

Figure 3: Item fit map for all items on the Australian adaptation of the Peabody Picture Vocabulary Test (PPVT-III) calibrated with all cases anchored to core items



'Who Am I?'

'Who Am I?' consisted of 11 pages on which children were to write their names, copy shapes and write words and numbers. Each response was assessed on a four-point scale relating to the skill required for the task. A score of zero was assigned if no attempt was made on the item. The data were fit using a partial credit item response model. The final case estimates were transformed to a scale with a mean of 64 and standard deviation of 8. Summary statistics are shown in Table 9 (on page 21).

In Wave 1, 4,827 children were administered 'Who Am I?' The test had a person separation reliability of 0.89.

Quality of the 'Who am I' test

The statistics indicate that the 'Who Am I?' data fit the rating scale model well, with most items falling within acceptable ranges, as shown in Figure 4. The most difficult item on the test was item 10, in which children were asked to write a sentence. Only nine children received four points for their response; more than one-half of children made no attempt on this item. This is also shown in the item fit map (Figure 4): item 10 (Sentence) has an infit mean square of 1.67, while all other items ranged from 0.77 to 1.14.

Figure 4: Item fit map for all items on the 'Who Am I?' test

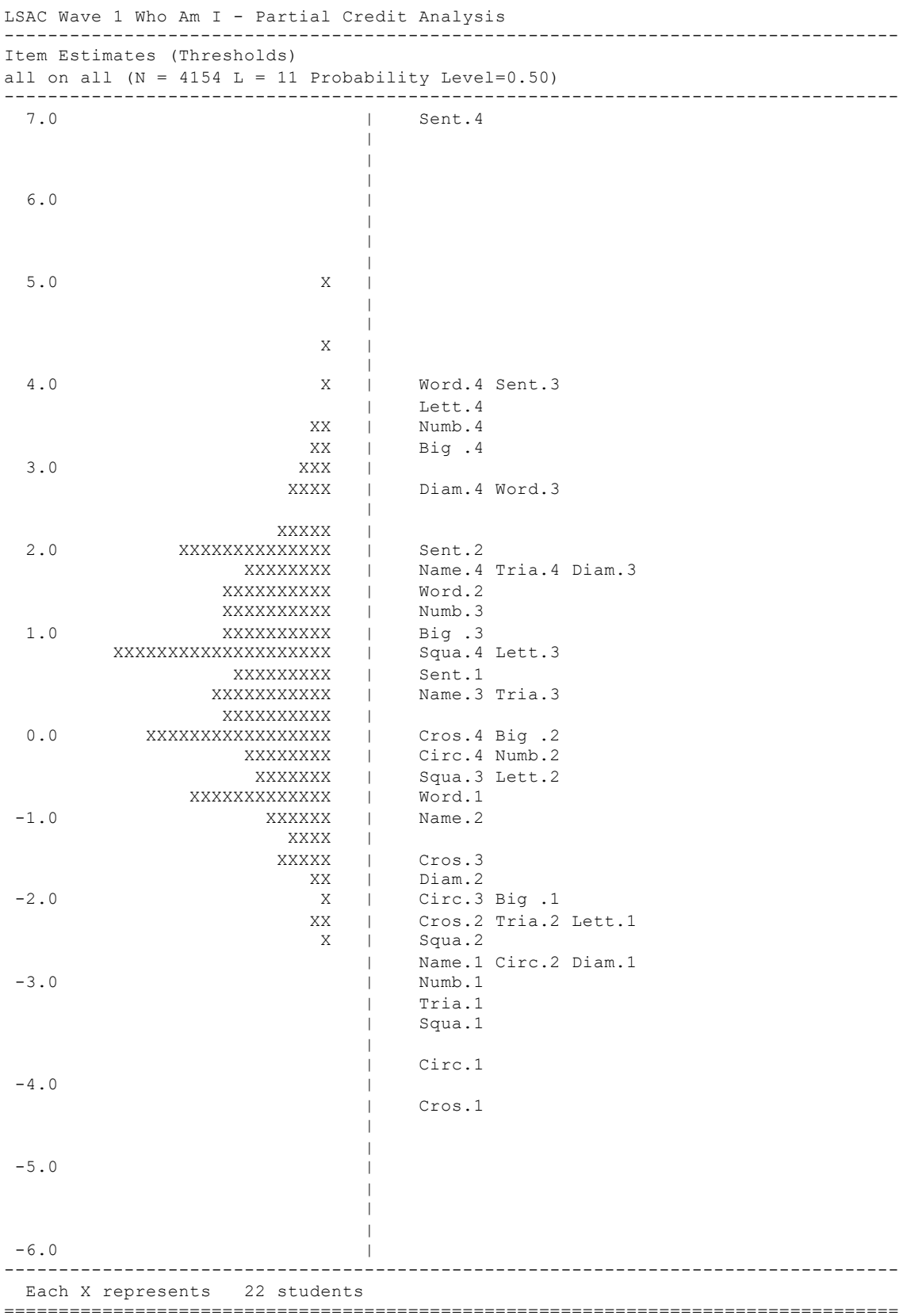
```
LSAC Wave 1 Who Am I   - Partial Credit Analysis
-------------------------------------------   ----------------------------------  -
Item Fit
all on all (N = 4154 L = 11 Probabilit   y Level=0.50)
-------------------------------------------   ----------------------------------  --
INFIT
 MNSQ             0.71      0.83      1.00      1.20         1.40      1.60      1.80
-------- ------- ----+--------- +--------- +--------- +-------  --+--------- +--------- +--
    1 Name                        *       |
    2 Circle                              |      *
    3 Cross                              *|
    4 Square                       *      |
    5 Triangle                     *      |
    6 Diamond                             |*
    7 Numbers                 *           |
    8 Letters              *              |
    9 Words                               |       *
   10 Sentence                            |                                  *
   11 Big John                  *         |
============================================================     =========== ====== ====
```

Analysis of a partial credit model provides information on the thresholds required to move from any score to a higher score on each item. This information is provided in Figure 5 (on page 24), the item map, and is plotted against the distribution of case estimates. For all items, higher scores had higher logit values, indicating that higher scores were achieved by children who had higher ability estimates. The item map also indicates that the distribution of children (on the left) was only slightly above the distribution of the items (on the right), indicating that there was a good match between the difficulty of the items and the children's ability levels.

Correlation between the adapted PPVT-III and 'Who Am I?'

The two tests had a Pearson product-moment correlation of 0.309, based on 4,386 children who received scores on both tests. While significant at the .01 level, this is considered a low-to-moderate correlation.

Figure 5: Item map for all cases on the 'Who Am I?' test



3.2 Wave 2 PPVT development

This paper describes the procedures used to develop a shortened version of the Peabody Picture Vocabulary Test (PPVT) for use in *Growing Up in Australia*: The Longitudinal Study of Australian Children (LSAC). This version of the test was developed to be used with six year olds as part of the second Wave of LSAC and is linked to the shortened version developed in 2003 for use with four year olds as part of the first Wave of LSAC (Rothman, 2013). The concept of a shortened version of the PPVT is from work done in the USA for the Head Start Impact Study. The original work was described in a draft paper by Philip Fletcher of Westat.

Procedures

All procedures described below are based on procedures used for the version used with four year olds. For that version, four alternative sets of items were tested; one set was developed for use in LSAC. For the six-year-old version, no alternative sets were used as the scope of the project was to develop a test that could be linked to the four-year-old version.

As done for the test administered to four year olds, the purpose was to develop a test that would consist of 40 items divided into a core set of 20 items, a basal set of 10 items for children who miss a minimum number of items on the core set, and a ceiling set of 10 items for children who correctly answer a minimum number of items on the core set. No child would take more than 30 items. It was also decided that at least 50% of children should be required to take the core set only.

Testing

A sample comprising 421 children was drawn from schools in New South Wales, Victoria and Queensland. During July and August 2005, the children were administered the full version of the PPVT-III, Form A, using the standard procedures for administering the test to six year olds. These children ranged in age from five years seven months to seven years 11 months. Seventy-eight per cent of the children were six years old, and 18% were seven years old. All children were in the same classes at the schools involved in the data collection. Subsequent examination of the data showed that the children from out-of-range ages did not appear as significantly different cases.

Analysis

Test items were examined using a one-parameter logistic IRT model with the software Quest. For items below the PPVT basal set that were not administered, all were marked as correct. Items with a low number of responses were eliminated from the IRT analysis. Overall, 132 items were used for analysis, as they covered a range that would allow 40 items to be selected and included the items administered in the four-year-old test.

Selection of items

The properties of the items were then determined, based on the data available from the Quest output. The first stage was to identify link items from the four-year-old test that could be used with the six year olds. For the 20 items of the core set, eight items that had appeared in the four-year-old test were selected. These items were selected on the basis of infit mean square and outfit mean square close to 1.00 in both administrations, the degree of difficulty on the items among both groups, the consistency of change between the administration to the groups, and the ability to provide a reasonable spread across the core set. Two items from the four-year-old test were selected for the basal set, and one item from the four-year-old test was selected for the ceiling set.

After the link items were selected, the remaining items were selected using those with infit mean square and outfit mean square close to 1.00, good discrimination and an ability to provide a reasonable range of item difficulties (–2.50 to +2.50). Items were also selected according to their position in the original PPVT sets and their parts of speech: nouns, verbs and adjectives. The final 20 core items were then positioned into two sets of 10 items, with the first 10 items generally easier than the second 10 items but with an overlap of item estimates. Similar analyses were done to select the 10 basal and 10 ceiling items.

Table 10: Items selected for adaptive PPVT-III for use with six year olds in LSAC

Set	PPVT-III Form A item number	Item	Item threshold	Infit mean square
Core 1	42	harp*	-2.55	1.01
	74	nostril*	-2.29	0.96
	56	furry*	-2.08	0.96
	52	diving*	-1.99	1.02
	78	horrified*	-1.44	0.99
	67	calculator	-0.38	1.10
	77	towing	-0.12	1.02
	91	clarinet	-0.02	1.07
	107	fern	0.53	1.03
	118	archery	0.88	0.98
Core 2	66	swamp*	-0.47	1.13
	90	interviewing*	-0.20	1.00
	96	vine*	0.10	0.97
	88	surprised	0.61	1.02
	68	signal	0.91	1.03
	114	injecting	0.97	0.99
	128	wailing	1.29	0.94
	131	foundation	1.85	0.98
	140	pastry	2.33	0.99
	125	valve	2.74	0.98
Basal	45	juggling	-4.98	0.74
	32	fountain*	-3.85	0.97
	40	farm*	-3.26	0.99
	47	tearing	-2.98	0.77
	49	parachute	-2.19	0.93
	71	vegetable	-1.70	1.04
	57	drilling	-1.62	0.92
	61	vehicle	-1.30	0.99
	75	vase	-1.21	0.94
	85	flamingo	-0.52	0.97
Ceiling	122	dilapidated*	1.11	0.98
	97	pedal	1.85	1.03
	149	abrasive	1.97	1.09
	143	pedestrian	2.07	0.97
	117	microscope	2.15	1.07
	153	detonation	2.69	0.94
	151	cascade	2.96	0.91
	139	consuming	3.57	1.04
	148	replenishing	4.58	1.14
	167	talon	- -	- -

Notes: Item threshold and infit mean square statistics are from the simulated test. *Link item included in test for four year olds.

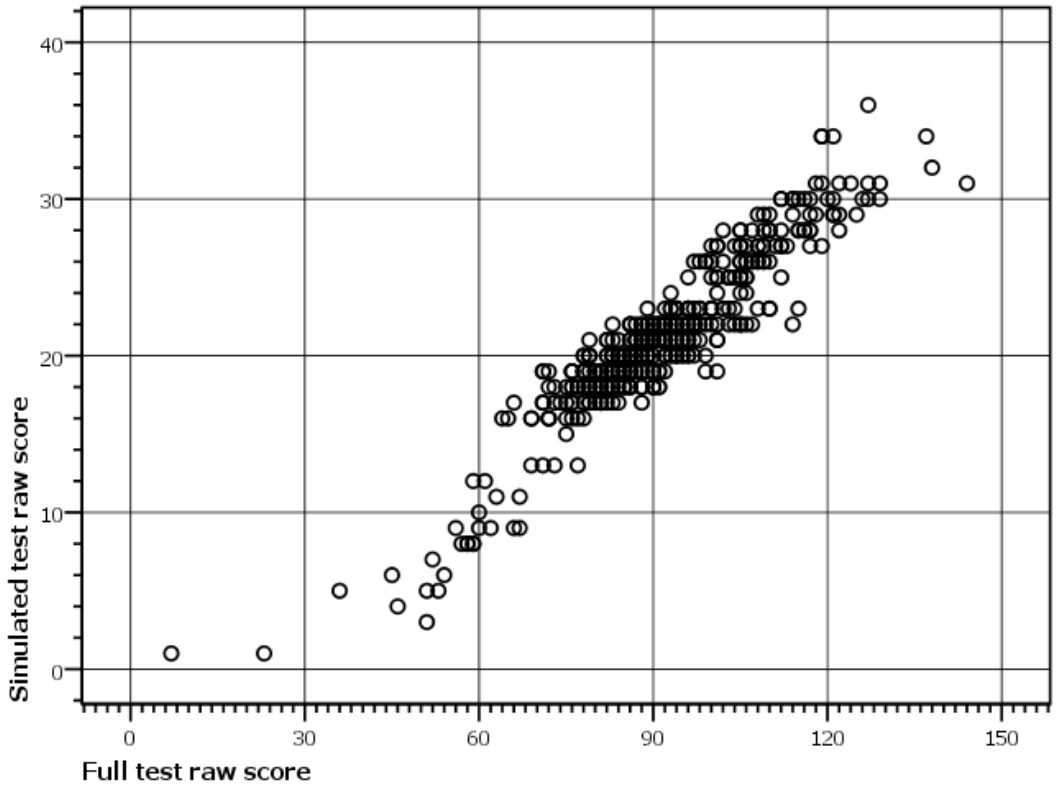
Simulations

Simulation results were then calculated in SPSS. With the objective of having approximately 50% of children requiring only the core set of 20 items, the items were checked to determine percentages of children requiring the basal or ceiling sets. Rules governing the administration of the test, particularly those regarding the number of correct items required for administration of the basal or ceiling sets, also guided the selection of items. The simulation suggested that approximately 25% of children would require the ceiling set, 10% the basal set and 65% the core only. It should be noted that similar targets for the four year olds were not achieved in the first Wave of LSAC, with more than 65% of children requiring the ceiling set.

Once the 40 items were selected, a new IRT analysis was conducted using only those items. Case and item estimates showed that the model fit the data well.

Results for the two versions – the full PPVT and the adapted version – were then compared: the full PPVT raw score with the case estimates from both the full test and the adapted test (Figure 6). The lowest correlation was 0.887, while most correlations were in the 0.93 to 0.97 range, suggesting that the adapted version of the PPVT provides similar results to the full PPVT.

Figure 6: Scatterplot showing joint distribution of scores on simulated adaptive PPVT-III and scores on full PPVT-III for six year olds



Comment

The six-year-old version of the adapted PPVT-III for use in LSAC was developed so that it can be linked with the four-year-old version of the test. This allows for the measurement of growth between administrations of the test. Forty items were selected for the test, with 20 items administered to all children as the core set (core sets 1 and 2). For children who made 15 to 20 errors, an additional basal set of 10 items would be administered and, for children who made 0 to 6 errors, an additional ceiling set of 10 items would be administered. Simulations showed that approximately two-thirds (67%) of children would require only the core sets, 7% would require the core and basal sets, and 26% would require the core and ceiling sets; this distribution was considered in the selection of items.

Acknowledgement

This chapter is largely based on the work of Australian Council for Educational Research.

4 Imputations to solve missing data problems in Wave 2.5

A number of the variables in the Wave 2.5 data files have higher levels of missingness than is usual for LSAC self-complete questionnaires. This chapter details the imputations made, and also those considered and rejected, in order to limit the amount of missing data. Using answers to other questions one could impute some of these 'missingness', and this was done wherever possible.

Examination of the questionnaires revealed the following main reasons for the high levels of missing data (note that many of these are not exclusive to Wave 2.5 but appear to be exacerbated by other problems with the Wave 2.5 questionnaire):

1. Formatting issues. On pages where questions were in two columns at the top of the page but then in only one column at the bottom of the page, some respondents missed the second column at the top of the page. This affected the following questions:

Cohort	Question no.	Description
B	2	TV/computer in other rooms
B	3	Electronic games system
K	2(a)	TV/computer in other rooms
K	3(a)	Electronic games system, mobile, iPod
K	22	Required to look for work
K	23	Partner working
K	24	Income

Note: Number of missings was exacerbated by the instruction that appeared under Q1 that said 'If you do not have a computer at home, go to Q6'.

2. Instructions to skip questions. There were a number of questions where the lead-in instructions requested that only people with particular characteristics (e.g. people who are currently working) complete the following question. It appears that some respondents may have skipped reading the preamble and made their own decisions about whether the question was relevant to them. Where this inconsistency led to people answering the question who should not have, their data was removed. However, missing data could not be replaced in most cases. This affected the following questions:

Cohort	Question no.	Description
B	23	Main reasons not in paid work
B	24	Plans about paid work
B	31	Effect of government benefits on attitudes to work
B	32	Attitudes to work
K	4	Computer use at home
K	5	Internet use at home
K	26	Effect of government benefits on attitudes to work
K	27	Effect of work on school involvement
K	28	Attitudes to work

3. 'None of the above' questions. A number of questions provided a 'none of the above' option if none of the other response categories applied. Experience with other self-complete forms has shown that it is not uncommon for people to omit ticking 'none of the above'. In general, it could be assumed that many of the responses to questions that had no response categories ticked are in fact 'none of the above'; however, some people may skip questions for reasons that are not readily apparent. This affected the following questions:

Cohort	Question no.	Description
B	2	TV/computer in other rooms
B	11	Child care
B	18	Life events
B	24	Plans for paid work
B	25	Government benefits
B	28	Current study, etc.
B	35	Child support arrangement services
B	43	Help from other parent
K	1	TV/computer in child's bedroom
K	2	TV/computer in other rooms
K	3	Electronic games system, mobile, iPod
K	11	Child care
K	16	Life events
K	21	Government pensions
K	31	Child support arrangement services
K	39	Help from other parent

4. 'Yes/No' questions. As with 'none of the above', it seems that some of the missing data for these questions could be explained by respondents for whom the 'no' response was relevant omitting to tick the 'no' option. In addition, if the question included a 'go to' instruction, then sometimes respondents forgot to tick whichever option they selected. The main questions affected are:

Cohort	Question no.	Description
B	16	Is study child the youngest?
B	19	Do you currently have a paid job?
B	33	Does the study child have a PLE?
K	14	Is study child the youngest?
K	17	Do you currently have a paid job?
K	20	Are you currently looking for work?
K	22	Are you required to look for work?
K	29	Does the study child have a PLE?

5. Questions where '0' is a valid response. These are often left blank as respondents feel that they don't apply. This affected Q14 for the B cohort (number of changes to child care arrangements), particularly since those currently without arrangements had been instructed to skip the previous two questions.
6. There were a number of cases in the B cohort (7) and the K cohort (29) that had roughly 90% of missing data. These cases have been excluded from the raw data files and the final files.

4.1 Rationale for imputations

Presence of media devices in the home and amount of time spent using these devices (B cohort Q2 to Q4)

An attempt was made to impute whether a child had access to the facilities listed in these questions by whether they had reported using them at Q4. Unlike many of the other imputations mentioned in this chapter, respondents were expected to answer the subsequent question regardless of their response to the previous ones. This meant some meaningful checks of the concordance between responses were possible.

However, most of the children (30 out of 41) who don't watch any TV still have TV in the home, so it couldn't be assumed that if children don't watch TV, they don't have access to one.

In Q1 and Q2 on the devices in their home, 46 of the 52 children were reported as watching some television in the home, even though these questions indicated that they did not have a TV. While this indicates a misinterpretation of at least one of the questions, it would appear that the presence of a television in the home couldn't be reliably inferred from a response indicating that the child watches television in the home.

So we can neither confirm nor refute the presence of a television in the home from the response to Q4a.

Likewise, 268 out of the 384 children that don't use a computer still indicated that they have one in the home, and 30 of the 51 children without a computer still use one. Again, the correspondence between the items isn't reliable enough to impute a response on whether there is a computer in the home.

For respondents with non-missing data for both the device ownership question (Q3) and the amount of time spent playing computer games (Q4), in only 25 of the 362 cases where the child doesn't have an electronic games system do they play one. Also, in 385 of the 435 cases where the child does have an electronic games system, do they spend some time playing it. Since this data follows the basic correspondence that would be expected, the presence or absence of a console has been imputed by whether the child plays with one when this information is missing. This has added 141 'no' responses and 107 'yes' responses. It has also been imputed that if the child doesn't have access to an electronic games system at home, the time spent playing with one at home will be nil, altering 1,541 responses from missing.

Devices in the home and possession of personal devices (K cohort Q2 and Q3)

No checks are possible on television use, electronic games systems or iPods. For mobile phones, there is no implication in the 'use of mobile phones' items at Q6 that the child has to use their own mobile phone to do these things, so ownership of a mobile phone can't be imputed.

For computers and the internet, cases where the child has a computer in their room are actually a little less likely to have one somewhere else in the home (87% vs 93%), so it can't be assumed that if they have one they'll have the other. Generally, if the respondent has given good answers to Q4 and Q5, they do have a computer in the home, but it's not necessarily the case that because the respondent has answered these questions incompletely that they don't have a computer. Therefore, it's impossible to impute accurately.

Presence or absence of child care (B cohort Q11, K cohort Q11)

If the respondent indicated that the child did spend time at child care, the child was assumed to have an 'other' type of child care for the K cohort; however, for the B cohort, they were imputed as having child care but the type of child care was set to missing since no 'other' option was available. This affected three cases in the B cohort and two cases in the K cohort. If the respondent reported zero for the number of days or hours per week of child care, it has been imputed that the child had no child care. This affected one case for the B cohort and two cases for the K cohort.

Is the study child the youngest child in the home (B cohort Q16, K cohort Q14)

The study child was aged either 3–4 years for the B cohort or 7–8 years for the K cohort at the time of the Wave 2.5 questionnaire. Therefore, if the respondent indicated that the age of their youngest child corresponds with this, it has been imputed that this is the study child, affecting six cases in the B cohort and 14 cases in the K cohort. If the age given was younger than this, it has been imputed that this wasn't the study child, affecting 14 cases in the B cohort and four cases in the K cohort.

Does the respondent have a paid job (B cohort Q19, K cohort Q17)

If the respondent indicated that they did work some hours then they were imputed to have a paid job, affecting two cases for the B cohort and 13 for the K cohort that were previously missing. If they said they generally work zero hours then they were imputed to have no job, affecting one case for the B cohort. If B cohort respondents were missing data for work hours and their desired number of work hours but had a response for why they were not currently working or for their future work plans, they were imputed as being out of work, affecting four cases for the B cohort. This question was not asked for the K cohort so no similar imputation was possible.

Some of the remaining missing cases had data for the desired work hours question; however, those with or without a job could logically answer this question, so this provided little indication of the true response to whether they were working. The attitude items for those in work (Q32 for the B cohort, Q27 and Q28 for the K cohort) could also be answered by some non-workers on the basis of previous work experience, so imputation based on responses to these was deemed unreliable.

Whether government benefits are received (B cohort Q25, K cohort Q21)

For the K cohort, if the respondent indicated that they are required to do an activity test, it could be imputed that a benefit is received; however, none of the missing cases met this criteria. The only other possibility for imputing this question would be to look at the effect of government benefits on work-plan items (B cohort Q31, K cohort Q26); however, there is no way of knowing if the respondents that didn't answer these questions were getting family tax benefits. Also, the skip is not very well highlighted in the formatting, so there can be little confidence that those who answered the question understood who it was for.

Whether the study child has a PLE (B cohort Q33, K cohort Q29)

A number of the missing cases have been classified on a case-by-case basis based on responses to the follow-up questions on child support. The criteria for these classifications involved the amount of missing data, the amount of data that might indicate the presence of a PLE (e.g. having a child support arrangement vs not having one), as well as whether a PLE was present at Wave 2. This created one extra 'yes' response and one extra 'no' response for the B cohort and four extra 'no' cases for the K cohort. After this process there were 16 cases that were missing all subsequent information for the B cohort and nine for the K cohort. Most children do not have a PLE; however, some of these cases could be from people who gave up on the questionnaire. It can be reasonably assumed that if they answered 50% of items in the most recent question required of them then they haven't given up on the survey and therefore are just cases without a PLE. This added an extra 12 'no' cases for the B cohort and three for the K cohort.

Respondent information (B cohort Q47)

Initially, there were 100 records that were missing the respondent information ('who completed this form?') in the B cohort file and 85 records in the K cohort file. ABS were able to correct 88 B cohort records and 69 K cohort records by matching the names of the people that completed the Wave 2.5 form to the names of people who participated in Wave 2. The location of this question may have been a factor in why there are missings, because the question was located at the bottom of the back of the form.

5 Review of main educational program of 4–5 year olds

5.1 K cohort

In investigating the quality of the data for the child's educational program type (*cpc06a4*) at Wave 1, concerns were raised in regard to the consistency of responses to this item with other information from the face-to-face interview and the teacher questionnaire. It was decided to provide a corrected version of *cpc06a4* as well as the original version. The correction involved two processes:

1. If teacher data was present and contradicted the value given by Parent 1, the value indicated by the teacher data was used instead.

   Or

2. If no teacher data was present, a number of checks were performed on the consistency of the parent's response with other data given (e.g. number of hours in care, the age of the study child, etc.). If a majority of cases with teacher data were corrected when they had the same combination of the original response and number of inconsistencies, then those without teacher data were corrected to the majority value. For example, it was found that among those cases with two or more inconsistencies whose original response was 'Pre-year 1 in a school', more than 50% of the teacher data, where available, indicated that the true response was 'Preschool in a school'. This value was therefore assumed to be most likely for these cases in the absence of teacher data.

More information on this process is provided in the Data User Guide (available from growingupinaustralia.gov.au/data-and-documentation/data-user-guide)

At Wave 3, respondents were asked to confirm the details of the educational program the child was in at the time of the Wave 2 interview two years prior, and were then asked about the details of the child's educational programs from ages three to up to six years prior (working backwards until either the child wasn't in an educational program or was in preschool/kindergarten).

Table 11 (on page 33) shows the information captured for each year. This section suggests improvements to the imputation based on this new data.

To determine how to best use these data, some determination has to be made as to their quality. As an initial check, the recall data were checked for reliability with themselves. The data were considered unreliable if there was a greater gap in year level than the number of years between time points, or a lesser gap unless there was an indication that a year level was repeated. This check revealed 14.5% of the cases were unreliable. The data collected at Wave 3 for these cases were not used to impute *cpc06a4*.

The data were then examined to quantify the number of inconsistencies with other data items from the Wave 1 questionnaire. The following circumstances were considered to be inconsistent:

- The child was in a 'pre-year 1 program' at school and was:
  - attending this program fewer than five days/week
  - attending this program less than 30 hours/week
  - younger than 55 months of age at Wave 1

  or
  - in 'Year 1' in Wave 2 unless indicated they had repeated a grade level.

- The child was attending a 'preschool' (other than in day care) and was:
  - attending this program for 30+ hours/week
  - more than 62 months of age at Wave 1

  or
  - in 'Year 2' at Wave 2.

Table 11: Variables capturing previous years educational programs for the K cohort at Wave 3

Questions	
1) What program did child attend the year before, that is in (3 years prior)?	
Year 1 (Grade 1)	→ 3
Pre-year 1 program	→ 3
Preschool/kindergarten program	→ 3
Long day care	→ 3
Home-schooled	→ 3
Other	→ 2
Child did not attend an educational program	→ End of recall
2) Other specify	→ Previous year
3) Was that located in a school?	
Yes	→ Previous year
No	→ epc59d?
4) Was it a ..?	
Preschool/kindergarten only centre	→ End of recall items
Preschool/kindergarten in a long day care centre	→ End of recall items
Mobile pre-school	→ End of recall items
Long day care centre	→ End of recall items
Other	→ End of recall items

Three different versions of this information were compared using these checks:

1. 'Original' – the original value entered from the Wave 1 face-to-face interview
2. 'Teacher' – the original data corrected when it disagreed with data obtained from the Wave 1 teacher questionnaire
3. 'Recall' – the information as recalled by the respondents at Wave 3 for four years prior.

Among those cases that had teacher data at Wave 1 and had reliable recall data at Wave 3,[2] 81% were found to have no inconsistencies when using the recall data. This compares with 65% with no inconsistencies using the original data and 84% when using the teacher data. So, it would seem that the teacher data is still the most consistent indicator of the true value; however, the recall data is also reasonably consistent.

In order to determine how to best use this data in the imputation, two different methods were tried. In the first, the recall data was substituted for the original data automatically. There was agreement between the value created using this scheme and the one using the teacher questionnaire data in 76% of cases.

For the second approach to imputation, the recall data (when reliable) were used as an additional check to those listed above and imputations were made on the basis of the number of unlikely combinations of data.

Under the second scheme, the following corrections were made:

1. Children in 'Year 1' were automatically recoded to 'pre-Year 1'.
2. Children in 'pre-year 1' with two or more inconsistencies were recoded to 'preschool in a school'.
3. Children attending a 'preschool in a school' with two or more inconsistencies were recoded to 'pre-year 1'.
4. Children attending a 'preschool at a non-school centre' with two or more inconsistencies were recoded as being in a 'day care centre with a preschool program'.

For cases with Wave 1 teacher data, the data generated by these corrections matched the teacher data in 80% of cases, better than using the recall data by themselves, and better than the correction scheme used prior to the recall data becoming available (which matched in 73% of cases). This approach has therefore been taken.

2 That is, minus the 14.5% mentioned above.

5.2 B cohort

Given the problems experienced for the K cohort at Wave 1, a different set of questions on educational programs was developed for the B cohort at Wave 3 (see Table 12). In Wave 1 for the K cohort, the data collected from the face-to-face interview on educational programs differed from that collected in the teacher questionnaire in 29% of cases. In Wave 3, for the B cohort, there were differences in 13% of cases.

However, when the consistency of the teacher data and the parent data was tested against other answers in the Wave 3 interview, it was found that neither version had many inconsistencies; however, the teacher corrected version had slightly more (3% versus 2.7%).

In the seven cases (so far) with inconsistencies when the teacher data were used, the teacher's response was 'pre-Year 1 school program' while the parent's was 'preschool program in a school'. These cases may represent programs that don't fall neatly into either category (e.g. classes at a pre-Year 1 level that children attend part-time), although there is no consistency in terms of state of residence of the children or the organisational basis of the school (e.g. independent versus state versus Catholic). Whatever the situation is with these cases, there seems to be little reason to correct the parent data or teacher data when there is little indication of which is correct.

Outcomes

1. *Teacher data still to be used to correct parent data when available in determining educational program at Wave 1 for the K cohort.*
2. *Recall data to be used as an extra consistency check within the existing process when imputing this information when teacher data is absent.*
3. *No imputation to be performed on Wave 3 B cohort educational program data.*

Table 12: Variables capturing current educational programs for the B cohort at Wave 3

Questions	
1) (Thinking about the arrangement the child uses for the most hours per week) is this located in a school?	
Yes No	→ 1 → 5
2) What class or program does child attend?	
Year 1 (Grade 1) Pre-year 1 program Preschool/kindergarten program Long day care Other, e.g. multi-age classes, early intervention	→ 4 → 4 → 4 → 4 → 3
3) Other specify	→ 4
4) Does child attend this program at	
A government school? A Catholic school? An independent or private school?	→ Further items → Further items → Further items
5) Which of the following best describes where child goes?	
Preschool/kindergarten only centre Preschool/kindergarten in a long day care centre Mobile preschool Long day care centre Other	→ Further items → Further items → Further items → Further items → 6
6) Other specify	→ Further items

6 Cleaning of income data

Following the original release of the data, users reported problems with outlying values in the continuous income variables (i.e. afn09a, afn09b, afn09m, afn09f, cfn09a, cfn09b, cfn09m, cfn09f). While this is not unusual for income, it appeared that some of these cases had unusual responses to other questions for those with such high incomes (e.g. more modest incomes reported when asked about combined yearly income at K20 of the face-to-face interview, more menial occupations). It appears that many of these are due to discrepancies between amount and time period when reporting income (e.g. giving yearly income as weekly). Many of these outliers have been subsequently cleaned up, although certain assumptions have been made to do so.

The process for cleaning the Wave 1 data used adaptations of the data query rules coded into the Wave 2 CAPI instrument. As well as providing a logical framework to underpin the investigation, this will also help in making the data more consistent longitudinally.

The rules used were as follows:

1. If a respondent's only source of income is government benefits or salary they should not report an income of $0 or a loss.
2. If profit or loss is a source of income then incomes >$200,000/year should be queried unless they also have a salary.
3. Where government benefits are the main source of income, incomes >$750 a week should be queried.
4. For all other combinations of income types, incomes >$260,000/year should be queried.

Cases identified by the first of these rules were all set to missing. Most of these seem to be due to respondents not counting government benefits as income. In the B cohort file there were 49 cases of this for Parent 1 and six for Parent 2, while for the K cohort, 31 cases were identified for Parent 1 and six cases were identified for Parent 2.

For those identified by the other three rules, if the categorical annual income for Parent 1 and Parent 2 at K20 was consistent with the continuous values, it was left as is. If there was an obvious correction that could be applied (e.g. deleting a zero from an income figure, changing the time period from weeks to year) to bring the income into or close to the range specified at K20 then this was applied. If there was no way that the continuous income values could be made reasonably consistent with the combined parental yearly income then the response was assumed to be an error and was made missing.

Restrictions on publishing case-level information limit what can be disclosed about these cases. However, for Parent 1s in the B cohort, of the 31 cases identified by rules 2–4, six were made missing, eight were corrected and 17 were left as is. For B cohort Parent 2s, of the 48 cases identified, four were made missing, 10 were corrected and 34 were left as is. For K cohort Parent 1s, of the 30 cases identified, five were made missing, six were corrected and 19 cases were left as is. For K cohort Parent 2s, of the 42 cases identified, five were made missing, eight were corrected and 29 were left as is.

In Waves 2 and 3, suspicious cases were identified using the above rules. These cases were checked against their income data from earlier waves, plus other information such as work hours and occupation. As would be expected, data collected with the CAPI instrument were cleaner, and fewer imputations had to be made. In Wave 2, seven corrections were made to Parent 1 income, four corrections to Parent 2 income, and one to the income of other adults in the home. In Wave 3, only one Parent 1 and one Parent 2 required correction.

7 Height differences

In the leave-behind questionnaires for both parents at Wave 1 and Wave 2, the parents were asked to report their height and weight so their body mass index (BMI) could be calculated. In cleaning Wave 2 data, it was discovered that there was a large number of discrepancies between the values reported by the same people at Wave 1 and at Wave 2. In fact, only 50% of respondents reported a value that was within 1% of their Wave 1 value.

Further investigation failed to find any explanation other than respondent error for the vast majority of these cases. In order that data analysts could assume that any observed changes in BMI were due to changes in reports of weight rather than height, it was decided to impute the value of height to be the average of the two reported values.

At Wave 3, the question on the Parent 1's height was asked of all new Parent 1's and those that had not returned a self-complete form at Wave 2, plus a handful of cases where Parent 1 had swapped places with Parent 2. However, for Parent 2, the height data was still collected by self-complete form, so sequencing cases around the question was not an option. Hence, for many,[3] there are now three points of data collection.

When the study child's height is measured as part of the interview process, a third measurement is taken if the first two disagree by more than 0.5 cm. If this is the case, the estimate of the child's height is considered to be the average of the two that correspond the most closely. This method of estimation means that the least reliable estimate has no effect on the result. It is suggested that in cases with three data points for a parent's height, the 'clean' result provided on the data file could similarly be the average of the closest two responses. As is done currently, the values of parental height for each Wave prior to this cleaning will remain on the data file if analysts wish to use their own approach.

Figure 7 shows the discrepancy between the two values used to create the 'clean' result for those parents with two data points versus those with three. Those with three data points had two that agreed in 77% of cases. Those with two data points had agreement in only 42% of cases. It should be noted, however, that at Wave 2, 45% of cases had agreement between the two data points, so there is some evidence that those who were more likely to return self-complete questionnaires were more likely to give accurate data.

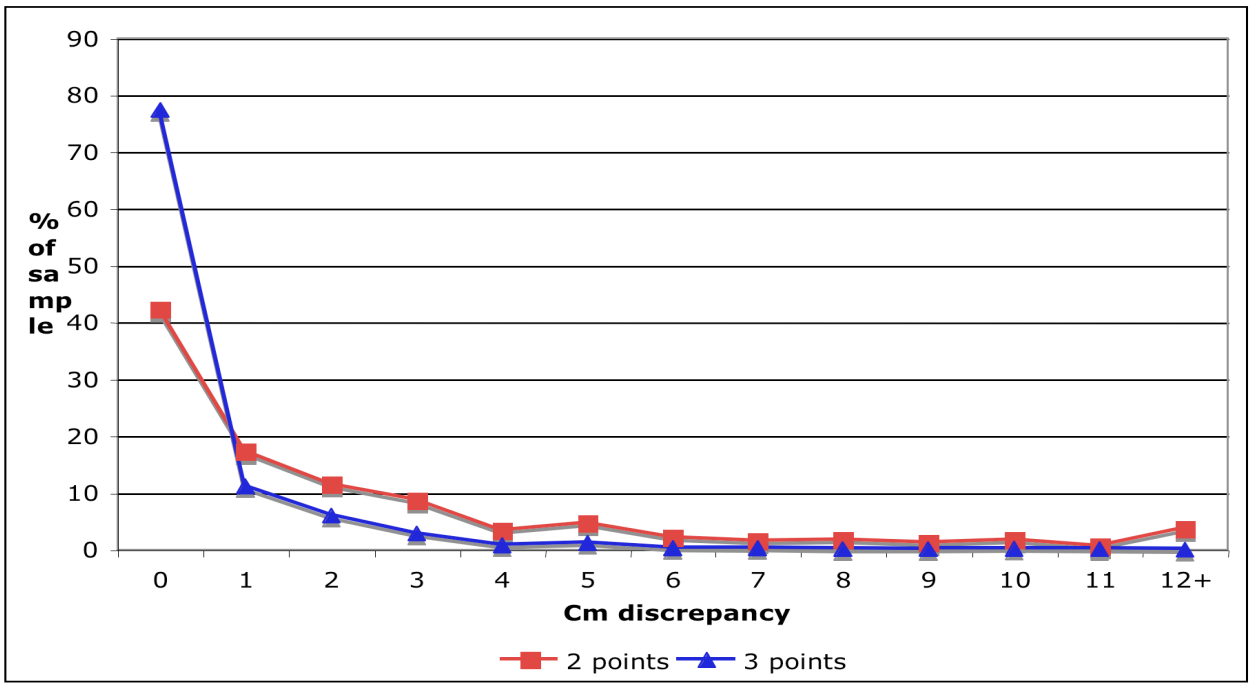
It was decided on inspection of Figure 7 (on page 36) that any case with more than a 10 cm discrepancy between the two closest values should be considered unreliable and therefore should be set to missing. This would affect 4% of Wave 3 parents with two data points and less than 0.1% of cases with three data points.

Outcome:

1. *This problem with the height data was presented to the February 2009 data Expert Reference Group Meeting and the group decided that if the differences are less than 10 cm we would average all three, otherwise we would average the closest two. Consequently, this is how the height data have been adjusted.*

3 72% of those who returned Parent 2 questionnaires at Wave 3 have data from all three waves. Of all Wave 3 Parent 2s, 11% had no height data, 16% had one data point, 22% had two data points and 51% had three data points.

Figure 7: Centimetre discrepancy in two closest data points for those with three vs two data points on parental height for Wave 3 respondents



8 Data issues in Wave 3.5

8.1 Q30B and Q24K

These questions ask how many hours per week the study child spends doing each of the following activities: watching TV, watching DVDs, using the computer, playing games, and listening to music. A total of 20 B and K respondents indicated an unusually large number of hours for some of these activities (that were not able to be corrected through normal editing processes). If the respondent indicated that the child spent more than 40 hours on weekdays doing an activity, these activities were set to missing.

This affected:

- nine cases for the B cohort and two cases for the K cohort for time spent watching TV on weekdays
- one K cohort case for time spent listening to music.

If the respondent indicated that the child spent more than 20 hours on a weekend doing an activity, these activities were also set to missing. This affected:

- six cases for both the B cohort and the K cohort in relation to time spent watching TV on the weekend
- two B cohort cases and three K cohort cases for time spent watching DVDs on the weekend
- two B cohort and eight K cohort cases for time spent using a computer on the weekend
- one B cohort and three K cohort cases for time spent playing a game console on the weekend
- one case for both the B cohort and the K cohort in relation to time spent listening to music on the weekend.

Care should be taken when using data on media use as there is no provision in the form for parents to report whether the activities were undertaken concurrently. For example, the TV may be on and the child may also be using the computer; therefore, it may be acceptable that some reports of total activities are greater than the total hours in a weekend. However, while watching TV and using a computer at the same time may be plausible, it cannot be clearly determined that this is what is occurring from the responses given.

8.2 Q13B and Q35K

These questions ask for the number of days per week during school term that the study child walks, rides, uses public transport or goes by car to and from school. Although respondents were asked to report the MAIN form of transport each day, some respondents reported multiple travel types.

The question aimed to obtain the main mode of transport used during an average school week to and from school. Therefore, the number of trips to and from school should sum to five each. However, this was not the case for some respondents, as shown in Table 13 (on page 39).

Table 13: Frequencies on Q13B and Q35K

	To school	From school
B cohort		
Less than 5	125	132
5	2,694	2,785
More than 5	126	116
Total	**2,945**	**2,901**
K cohort		
Less than 5	35	77
5	2,739	2,674
More than 5	183	181
Total	**2,957**	**2,932**

The cases in which the number of trips to or from school did not add to five have been set to missing. This affected 296 in the B cohort and 319 in the K cohort in total. Two hundred and three B cohort and 157 K cohort cases had this problem for both to and from school. Forty-eight B cohort and 61 K cohort cases had a problem with to school, while 45 in the B cohort and 101 in the K cohort had a problem with from school.

8.3 Q25B/Q26B and Q12K/Q13K

Instructions to skip questions

There were a number of questions where respondents did not follow the skip instructions correctly. Where this inconsistency led to people answering the question who should not have, their data was removed. However, missing data could not be replaced in most cases. This affected the following questions:

Cohort	Question no.	Description
B	5	Looked forward to going to school
B	6	Upset or reluctant to go to school
B	8	Teacher informs parent of child's progress behaviour
B	9	How well child gets along with other children
B	10	Quality of education
B	11	Teacher/Parent relationship
B	12	School social capital
B	13	Travel to and from school
B	14	Distance to school
B	15	Provide help with homework
B	16	Experiences before child started school
B	17	Communication with school before child started
B	18	Child's transition to school
K	20	Child began to menstruate

‘None of the above’ questions

A number of questions provided a ‘none of the above’ option if none of the other response categories applied. Experience with other self-complete forms shows that it is not uncommon for people to omit ticking ‘none of the above’. In general, it could be assumed that many of the responses to questions that had no response categories ticked are in fact ‘none of the above’. However, some people may skip questions for reasons that are not readily apparent. In these cases, the item has been left as missing. This affected the following questions:

Cohort	Question no.	Description
B	16	Experiences before child started school
B	17	Communication with school before child started
B	27	Sleep problems
B	28	Stressful life events
K	21	Stressful life events

‘Yes/No’ questions

As for ‘none of the above’, it seems that some of the missing data for these questions could be explained by respondents for whom the ‘no’ response was relevant, omitting to tick the ‘no’ option. The main questions affected are:

Cohort	Question no.	Description
B	33	Family rules
K	6	Family rules about homework
K	7	Special place to do homework
K	28	Family rules

Substantial amounts of missing data

There were some cases in the B cohort (one) and the K cohort (five) that had roughly 90% or more missing data. These cases have been excluded from the raw data files and the final files.

9 Data issues in Wave 4

9.1 Instances of child but not parent participation

Ideally, it is expected that parents who provide consent to interview their children would agree to be interviewed themselves. However, this does not happen in 100% of cases. In Wave 3, there was only one case out of 8,718 home visits in which a parent was not interviewed but provided consent to interview a child. In Wave 4, there were five out of 8,405 cases where parents were not interviewed but agreed to their children being interviewed. The main reasons for parents to refuse a home interview but allow their child to be interviewed were lack of interest and time.

Parent 1's refusal to the home interview might lead to missing household information.[4] If household information is not available this record is not included in the household file. As a result, in Wave 4, the household file hhgrk10 contains 4,164 records and main file lsacgrk10 contains 4,169 records. If data users intend to merge these datasets, they need to be aware that there is a mismatch between the datasets for five cases.

To help a data user identify cases with available household data the following variables were created: *hhresp for both cohorts. In variable names the asterisk refers to an age indicator (In Wave 4, d refers to the B cohort and f refers to the K cohort).

While in Waves 3 and 4 the discrepancy in child and parent participation is minor, it might increase in future waves due to: (1) changes to the interview procedure; or (2) children becoming more active participants/refusers in the study; or (3) increases in other activities of parents and children meaning fewer times when both are at home at the same time and available for interview.

Changes to the interview procedure

From Wave 4, interviewers were provided with two laptops and were able to conduct 'parallel interviewing'. The interview was split into two streams; all the Parent 1 (P1) questions were on laptop 1 and all the study child (SC) questions were on laptop 2. As a result, the interviewer had the flexibility to complete child and parent interviews either at the same time or at different times.

Children becoming active participants/refusers of the study

All cases where children were interviewed but parents were not belong to the K cohort in both Waves 3 and 4.

9.2 ACASI

The audio computer-assisted self-interview (ACASI) contains questions that are skipped if the study child has no mother and/or father figure in their life or does not attend school. As these circumstances were determined in the CATI component of the Parent 1 interview on laptop 1, they were not apparent in the interview on laptop 2 when the ACASI was conducted. Consequently, in order for the relevant questions to be skipped, prior to providing the laptop 2 to a study child to complete the ACASI, the interviewer was required to enter into the ACASI instrument whether there was 'Mum' and/or 'Dad' figures in the study child's life and whether the study child had been attending school. The interviewers were asked to use their knowledge of the family that they gained after completing the CATI component with the Parent 1.

4 If a Parent 1 completes CATI prior to the home interview, household information is not missing.

When deciding if it would be appropriate to ask about a 'Mum' or 'Dad' the interviewers were asked to be sensitive towards the situation of the child as family structure can be complicated. The interviewers were instructed that if they were unsure whether there was a 'Mum' or 'Dad' in the child's life and/or whether a child had any contact with them to enter 'no father' or 'no mother' so as to not distress the child. For example, in the situation where the current Parent 1 and Parent 2 were not the biological parents, it was unclear as to who the child would be referring when asked about their mother or father. In just a few cases there was an interviewer's error and wrong information was rolled into the ACASI module.

In just a few cases there were inconsistencies between the household information and the interviewer's assessment of whether there was a mother in the child's life. As a result, there were 10 cases for which a mother was recorded in the household file but where all questions about the mother were skipped in the ACASI module. There were also 32 cases where all questions about the father in the ACASI module were skipped but a father was identified in the study child's life in the household file. To identify these problematic cases the following variables were created (asterisk refers to an age indicator):

- *mumsk – a mother is identified in the household but questions about 'Mum' figure are skipped in ACASI module
- *dadsk – a father is identified in the household but questions about 'Dad" figure are skipped in ACASI module
- *schsk – child is in school but questions about school are skipped in ACASI module.

With regards to school attendance, 40 children who were in school (as identified in the education module) skipped all the questions about school in the ACASI module. This mismatch was mainly due to manual errors and one of the problems of the method used; that is, information about the child's education is provided by Parent 1 and stored on the laptop 1 and the study child completes ACASI separately on the laptop 2. This information being stored on different computers means that these instruments do not talk to each other.

9.3 Matrix Reasoning

Matrix Reasoning (MR) is a test from the Wechsler Intelligence Scale for Children, 4th edition (WISC-IV) (Wechsler, 2004) for ages 6–7, 8–9 and 10–11 years. This test of non-verbal intelligence presents a child with an incomplete set (later referred to as item) of pictures and requires the child to select the picture that completes the set from five different options. The instrument is comprised of 35 items of increasing difficulty.

Administration rules

According to the WISC-IV manual, the administration of Matrix Reasoning should follow a set of rules. We are not going to discuss all the rules in detail but rather focus on the rules crucial for our purposes.

Administration of the test should start at the age-specific start-point, which is indicated in the WISC-IV manual. Item 4 is the start-point for children aged 6–7 (B cohort) and Item 7 is the start-point for children aged 10–11 (K cohort).

Items prior to age-appropriate start-points are called reversal items. Reversal items are asked only if a child provides incorrect answers on the first or second start-point item. If a child answers incorrectly either of the first two items from the start-point, the interviewer asks the preceding items (reversal items) in reverse sequence until the child answers correctly two consecutive items and then goes back to the age-appropriate items and proceeds with the rest of the test. This is called reverse administration. For example, if a 6-year-old child answered correctly on Item 4 and incorrectly on Item 5, an interviewer should reverse to Item 3, then Item 2. If Item 3 or Item 2 is incorrect then Item 1 is administered. If Items 3 and 2 are correct, Item 1 is not administered. After administering reversal items, the interviewer goes back to Item 6 and proceeds with the rest of the test.

Scoring rules

The total raw score of MR is equal to the number of correct items starting from an age-appropriate start-point plus the total score on the reversal items. For items administered from the age-appropriate start-point a raw score of 1 is assigned for each correct answer.

For reversal items the following scoring rules are applied:

- Rule 1 – Each reversal item gets a score of 1 if the reverse administration is not required (first two items from the start-point are answered correctly). For example, if a 6-year-old child answers correctly Items 4 and 5, the reversal Items 1, 2 and 3 are scored 1 each.
- Rule 2 – Each reversal item gets a score of 1 if a child correctly answers two consecutive reversal items. For example, if a 6-year-old child answers correctly on reversal Items 3 and 2 and Item 1 is not administered or answers incorrectly on Item 3 and correctly on Items 2 and 1, the reversal Items 1, 2 and 3 are scored 1 each.

- Rule 3 – Each correctly answered reversal item gets a score of 1 and each incorrectly answered reversal item gets a score of 0 if a child does not answer correctly on any two consecutive reversal items. For example, a 6-year-old child answers incorrectly on the reversal Items 3 and 1 but correctly on the reversal Item 2. Then, Items 3 and 1 are scored as 1 each and Item 2 is scored as 0.

Administration of MR in LSAC

Due to the technical difficulties in programming, the reverse administration was not implemented in the LSAC MR instrument; that is, if LSAC children answered either of the two items from the start-point incorrectly the reversal items were never administered. Table 14 shows a number of cases where the first two items from an age appropriate start-point were answered correctly and incorrectly for B and K cohorts.

Table 14: Frequencies of correct responses on the start-point items

	n	%
B cohort		
Item 4 and Item 5 are correct	3,964	95
Item 4 or Item 5 is incorrect	216	5
Total	**4,180**	**100**
K cohort		
Item 7 and Item 8 are correct	3,908	95
Item 7 or Item 8 is incorrect	195	5
Total	**4,103**	**100**

It can be seen from Table 14 that 95% of children answered the first two items from the age-appropriate start-point correctly and did not require the reverse administration. The raw scoring for these children was based on Rule 1. The other 5% of children answered one of the first two administered items incorrectly and, therefore, required the reverse administration to identify which rule for scoring should be used – Rule 2 or Rule 3. Given that the reverse administration was not available, it was decided to assign all reversal items a raw score of 1 regardless of whether the first two administered items were answered correctly or not. As a consequence, some of the 5% of children might have had their MR scores overestimated. The following variable was created to identify these 216 cases in the B cohort and 195 cases in the K cohort:

$$*mrrawi = \begin{cases} 1, \text{ if either of one first two items from start point is incorrect} \\ 0, \text{ otherwise} \end{cases}$$

where * refers to appropriate age indicator.

The MR scores on Items 1–6 from previous waves are examined below.

K cohort

Out of 195 children from the K cohort who did not answer either one of two first items from the start-point at Wave 4, 185 children did MR at Wave 3. While the reverse administration was not implemented in Wave 3, all items were administered; that is, Item 1 was the first administered item. This allows us to cross-check how many children out of 185 gave two consecutive correct answers on Items 1, 2, 3, 4, 5 and 6. There were 179 who answered correctly either on all Items 1, 2, 3, 4, 5 and 6 or answered correctly on two consecutive items. In this instance, at Wave 3, they were assigned the maximum possible score. Assuming that cognitive ability of children remains stable over time, we would expect these children would obtain the maximum possible score for the first six items at Wave 4 too.

B cohort

In Wave 4, B cohort children were administered the MR test for the first time. However, in Wave 2, K cohort children did the MR test and they were the same age as the B cohort children in Wave 4. Therefore, the relative comparison could be made against the K cohort children of the same age. In Wave 2, there were 269 (6%) of the K cohort children who answered Items 4 or/and 5 incorrectly. Out of these 269 children, only 16 children did not answer correctly Items 3 and 2 or Items 2 and 1.

Therefore, based on the data from previous waves, we would expect only a very small number of children in either cohort to have their MR ability overestimated through the changes in administration and scoring.

10 Data issues in Wave 5

10.1 Geography

The first four waves of LSAC data included geography items such as postcodes and various levels of the Australian Standard Geographical Classification (ASGC) that were generated from geocoding of the residential addresses of study families. In Wave 1, the geocodes were based on global positioning system (GPS) coordinates obtained by I-view interviewers at the time of interview, while in Waves 2–4, they were based on residential addresses collected by Australian Bureau of Statistics (ABS) interviewers.

In July 2011, the ABS introduced a new statistical geography framework called the Australian Statistical Geography Standard (ASGS) to replace the ASGC. The main purpose of the ASGS is to disseminate geographically classified statistics. It provides a common framework of statistical geography enabling the publication of statistics that are comparable and spatially integrated.

Improved data sources and technology have allowed the ABS the opportunity to create a better geography optimised for the release of ABS statistics. A new robust and stable structure means that changes over time are minimised, assisting in the maintenance of quality time-series data. In addition, the ASGS, together with improved methods of calculation, allows for more accurate correspondences to translate ABS data to non-ABS administrative and geographic regions.

For further information on this new standard refer to 1270.0.55.001 – *Australian Statistical Geography Standard (ASGS): Volume 1 – Main Structure and Greater Capital City Statistical Areas, July 2011*.

To take advantage of this more comprehensive, flexible and consistent way of defining Australia's statistical geography, the ASGS is included on LSAC releases from Wave 5 onwards. To ensure that there is a common geographical standard across waves, the decision was made to:

- dual-code Wave 5 residential addresses to ASGC and ASGS, enabling the comparison of old and new classifications; and
- back-code Waves 1–4 residential addresses to the new standard ASGS.

The new variables added to the general release file for each Wave are shown in Table 15.

Table 15: New geography variables included from Wave 5

Without age variable name	Label
Gccsa	Australian Statistical Geography Standard (ASGS) – Edition 2011 – Greater Capital City Statistical Area Structure
Sos	Australian Statistical Geography Standard (ASGS) – Edition 2011 – Section of State
sa22011	Australian Statistical Geography Standard (ASGS) – Edition 2011 – SA2
sa32011	Australian Statistical Geography Standard (ASGS) – Edition 2011 – SA3
sa42011	Australian Statistical Geography Standard (ASGS) – Edition 2011 – SA4
absra	Australian Statistical Geography Standard (ASGS) – Edition 2011 – Remoteness Area (ABS)

Most addresses were auto-coded using ASGS address coders, which link addresses to geographical areas. However, in some cases, addresses were either incomplete, had spelling errors or, more rarely, were identical addresses in the same suburb. In these cases, addresses were manually cleaned to reduce the number of records with missing geocodes. After these steps, there were still some records unable to be geocoded to ASGS (level SA2). These numbers for Waves 1–5 are provided in Table 16.

Table 16: Number of records missing SA2 by wave

Wave	Number of responding records not coded to SA2
1	20
2	12
3	13
4	34
5	2

To enable coding to the ASGS, many addresses needed cleaning to ensure accurate data. As a result, some records have SLAs where there were none previously, and others have been coded to a different SLA.

The 2011 Census and SEIFA data are available in the new ASGS classifications. However, while it is possible to provide ASGS classifications for Waves 1–5, census and SEIFA data for 2001 and 2006 are not available for these new geographic classifications (ASGS).

From Wave 7 onwards only ASGS geography variables will be output on the files.

10.2 Occupation

LSAC data include variables for the occupation of Parent 1 (P1) and Parent 2 (P2). In recent waves, the occupation of Parent Living Elsewhere (PLE) and the parents of P1/P2/PLE (i.e. the study child's grandparents) are also included. These were coded using the Australian Standard Classification of Occupations (ASCO). The ANU4 scale – a scale of occupational status calculated using ASCO, which is an occupational classification system that classifies jobs according to skill level and skill specialisation – is also provided to data users for Waves 1–4.

Since Wave 2, LSAC occupation data have also been coded to the newer occupation standard, which is the Australian and New Zealand Standard Classification of Occupations (ANZSCO). ANZSCO was introduced in 2006 and was a product of a development program between the ABS, Statistics New Zealand and the Australian Government Department of Employment and Workplace Relations.

For further information on this standard, refer to 1220.0 – ANZSCO – *Australian and New Zealand Standard Classification of Occupations, First Edition*, 2006.

The latest release of ASCO was in 1997, reducing its applicability to the current Australian workforce. Therefore, from Wave 5 onwards only, ANZSCO codes will be produced. To enable the transition to using ANZSCO, the study has:

- added ANZSCO codes to the Waves 2–4 data files, as these codes were already generated during these waves, and is investigating the possibility of providing ANZSCO for Wave 1 through correction code
- replaced the ANU4 scale from Wave 5 onwards with the Australian Socioeconomic Index 2006 (AUSEI06) (McMillan, Beavis, & Jones, 2009), the latest in the series of occupation status scales developed by the ANU.
- provided AUSEI06 for Waves 2–4, and is investigating the possibility of adding to Wave 1 through a correction code.

The new variables added to the general release file are in Table 17.

Table 17: New occupation variables included from Wave 5

Question ID	Label
pw08_5	Current occupation (ANZSCO code)
pw08_6	Current or most recent occupation (ANZSCO code)
pw08_7	Current occupation (AUSIE06 code)

The SEP variable (Z score for socio-economic position among all LSAC families) has been calculated from Waves 1 to 4 using ASCO classifications. Due to ASCO being unavailable for Wave 5, the SEP variable has not been calculated and hence is not available in the Wave 5 dataset. Further work will be done into ways we can calculate the SEP using the ANZSCO classifications and a new/revised SEP variable may be available in the future.

10.3 ACIR data issue (all waves)

After analysis of the ACIR data previously supplied, it came to light that immunisation rates in LSAC did not reflect national rates. After investigation with the data provider, it was found that data extraction up to Wave 5 had not extracted all the required records. These data have been rectified; however, data users should not use the previous version of the ACIR data.

10.4 Changes to household files

Addition of 'Person Type' to the files

In Wave 5, Person Type (f21a) is available on the Waves 1–5 files for the first time, with a code attached to each household member and wave. This item is derived from information collected in the P1 interview and amended where needed during processing. A list of the person types and a description of each is shown in Table 18.

Table 18: Person Type descriptors

Code	Person Type	Description
1	Study child	The study children are the focus of the study, and consist of two cohorts (B cohort aged 8–9 years and K cohort aged 12–13 years in Wave 5).
2	Parent 1	Parent or guardian who provides the greatest role in caring for the study child and is therefore likely to be the most reliable informant on the health, development and care of the study child. Parent 1 must live with the study child
3	Parent 2	Study child's other resident parent/guardian, or the married or de facto partner of Parent 1. Another person in the household can be considered as Parent 2 if they are acting as a significant parental figure who helps to care for the child and is a stable member of the child's residential family unit.
4	Usual resident	A person other than the study child and the study child's resident parent(s) who usually lives in the study child's house (e.g. siblings of the study child)
5	Non-resident	A person other than a parent who has previously been a resident of the household but no longer lives in the same household as the study child
6	Parent living elsewhere	A parent of the study child who does not live in the same household as Parent 1 and the study child. This person may previously have been a Parent 2 (or a Parent 1).
7	Temporary member	Includes people who, in-between waves, joined the study child's household for more than three months but have since left
8	Empty row	In the household files row/member number 3 is always used for Parent 2 at Wave 1. When there was no P2 in the house at Wave 1, this row is left as an empty row. Also used when duplicate members are picked up
9	Deceased	A person who was previously recorded as a resident of the household but has died

Changes in relationship to study child information for household members

For Waves 1–4, the household file carried forward the relationship to study child for each member in the household from Wave 1 or the subsequent Wave for members entering the household after Wave 1. This means that for an existing household member, the relationship information in the household file is generally the same across waves. In some cases, this will not reflect changes in the relationships within the household. Relationship changes that we know did occur include:

- a step-parent changing to adopted parent
- an unrelated adult changing to step-parent
- a foster sibling changing to adopted sibling.

From the Wave 5 interview onwards the relationship of existing household members to the study child can be updated during the interview for household members present in previous waves.

As a result, from Wave 5 onwards there will be differences in the relationships between study children and household members between waves.

Inclusion of two waves of household data in the PLE person grid

The person grid is a list of people and their demographics associated with the study child, some members may still reside with the study child and others may have left. The Wave 5 parent living elsewhere survey instrument included roll-forward person grid data from Wave 4, so now two waves of household data for ongoing responding PLEs are available. Including Wave 4 details of a PLE's household in the survey instrument enables comparisons of the PLE's household circumstances between waves.

Concordance between people on main and PLE person grids

The concordance between the main household and the PLE's household has been provided for the first time in Wave 5. This enables the identification of who is the same person between the two files, who is on the main file only, and who is on the PLE file only. Table 19 provides a list of variables provided in the concordance file.

Table 19: Concordance file variables

Question ID	Label
MID5	Wave 5 Main Household Member Number
PLEID5	Wave 5 PLE Household Member Number
HHTYPE 5	Wave 5 Household Type
CHHFLOOP	Wave 5 Combined Household Row Number

The values for HHTYPE_5 are:

- 0 = Not present at Wave 5
- 1 = Wave 5 main household member only
- 2 = Wave 5 PLE household member only
- 3 = Wave 5 main and PLE household member

For example:

- Main household member number 4 was present at Wave 5, and that person was also present at Wave 5 in the PLE household, where they were recorded as member number 3. The variables that link these records will contain the following values: MID5 = 4; PLEID5 = 3, HHTYPE_5 = 3.
- If main household member number 4 was in the main household only at Wave 5, the values would be: MID5 = 4; PLEID5 = –9, HHTYPE_5 = 1.
- If PLE household member number 3 was in the PLE household only at Wave 5, the values would be: MID5 = –9; PLEID5 = 3, HHTYPE_5 = 2.

The values in MID5 and PLEID5 correspond to the member number in the data files, so this will enable you to find demographic information and link it to the files if required.

Child report of whether at school

At the start of both the study child's audio-computer-assisted self-interview (ACASI) module and the face-to-face Child Self-Report K (CSRK) module, the interviewer records whether the study child is attending school, using response options of Yes and No. If the study child does not attend a school, some questions about schooling are not asked. These questions are directly related to the school environment and therefore are not relevant to study children not attending school. Parent 1 is also asked a question about whether the child:

- attends a government school
- attends a Catholic school
- attends an independent or private school
- is not in school.

In total, the number of K cohort children coded as not in school as a result of the P1 interview was 33, whereas from the child interview the combined number was 218. Table 20 demonstrates that there were 191 records where the responses about whether the child was in school conflicted between the two interview components.

Table 20: Whether in school according to Parent 1 and study child components

Parent 1 (EDUC14)	Study child (ACASI02/CSRK02)				
	In school	Not at school (either question)	No study child interview	Neither question answered	Total
In school	3,639	189	51	38	**3,917**
Not at school	2	29	2	0	**33**
No P1 interview	4	0	0	0	**4**
Question not answered	1	0	1	0	**2**
Total	**3,646**	**218**	**54**	**38**	**3,956**

Table 21 cross-tabulates possible reasons for the discrepancy against school type, as recorded in the P1 interview for these 189 records. Around 44% of the difference seems to be accounted for by the interview taking place at the weekends or in school holidays.

Table 21: Characteristics of child or interview for children entered as not attending school by the interviewers

School attended	Interview date in school holidays	Interview date on weekend (not school holidays)	Interview date is school day	Total
Government school	32	16	61	**109**
Catholic school	11	5	23	**39**
Independent or private school	12	7	22	**41**
Total	**55**	**28**	**106**	**189**

To improve the quality of reporting in Wave 6, and to clear up any confusion, school attendance was recorded in the same way in both the child interview and the Parent 1 interview. In the child interview the same response categories of government school, Catholic school, independent or private school, and not in school will be provided instead of Yes/No responses. This change is to make it clearer that the study is asking about usual school attendance and not whether school was attended on the current interview date. This point was further highlighted in interviewer training.

11 Smoking inside the household

In Wave 3 there was a higher number of families recorded as having five or more people who smoked inside the household than in other waves (see Tables 22 and 23 below).

Table 22: Number of residents who smoke inside – B cohort

	No. residents smoke inside									
B cohort	**Wave 1**	**%**	**Wave 3**	**%**	**Wave 4**	**%**	**Wave 5**	**%**	**Wave 6**	**%**
Refused (-3)	0	0.0	0	0.0	16	0.4	8	0.2	2	0.1
Not answered (-9)	766	15.0	0	0.0	184	4.3	251	6.1	217	5.8
Missing (.)	40	0.8	1	0.0	0	0.0	0	0.0	0	0.0
0	3,815	74.7	4,060	92.6	3,759	88.6	3,485	85.3	3,276	87.0
1	318	6.2	136	3.1	205	4.8	228	5.6	174	4.6
2	141	2.8	58	1.3	68	1.6	97	2.4	86	2.3
3	18	0.4	3	0.1	6	0.1	13	0.3	6	0.2
4	5	0.1	1	0.0	1	0.0	1	0.0	1	0.0
5 or more	4	0.1	127	2.9	3	0.1	2	0.1	2	0.1
Total	**5,107**	**100.0**	**4,386**	**100.0**	**4,242**	**100.0**	**4,085**	**100.0**	**3,764**	**100.0**

Table 23: Number of residents who smoke inside – K cohort

	No. residents smoke inside									
K cohort	**Wave 1**	**%**	**Wave 3**	**%**	**Wave 4**	**%**	**Wave 5**	**%**	**Wave 6**	**%**
Refused (-3)	0	0.0	0	0.0	15	0.4	19	0.5	2	0.1
Not applicable (-9)	754	15.3	0	0.0	196	4.7	238	6.0	247	7.0
Missing (.)	54	1.1	1	0.0	0	0.0	0	0.0	4	0.1
0	3,631	73.7	3,939	90.9	3,652	87.6	3,366	85.1	2,994	84.6
1	399	8.1	197	4.5	201	4.8	199	5.0	182	5.1
2	123	2.5	78	1.8	88	2.1	116	2.9	90	2.5
3	15	0.3	3	0.1	10	0.2	12	0.3	9	0.3
4	5	0.1	4	0.1	4	0.1	3	0.1	6	0.2
5 or more	2	0.0	110	2.5	3	0.1	3	0.1	3	0.1
Total	**4,983**	**100.0**	**4,332**	**100.0**	**4,169**	**100.0**	**3,956**	**100.0**	**3,537**	**100.0**

This difference in response is likely to be due to incorrect recording of 'none' responses as '5' in the instrument, as '5' is the standard way for interviewers to record a 'no' response. In Wave 1 this item was collected as part of the Parent 1 leave-behind form and in Wave 3 this question was collected by the interviewer in a face-to-face interview. However, from Wave 4 onwards this question was changed to a computer-assisted self-interview (CASI), which the respondent completes themselves and, as a result, interviewer reporting error was not an issue. The change in collection mode may have also affected the responses to other categories if there was response

bias due to reporting smoking behaviours in a face-to-face interview rather than within the CASI, which is completed alone by the respondent.

To correct this issue, responses to other waves and the number of people in the household were used to either amend the responses or set them to missing where it was unclear what the response should be. Where reported responses to other waves were none, the Wave 3 data were set to none. If other responses were reported in other waves the data was set to missing, with the exception of two cases that reported four or five people smoking in other waves. This resulted in the following changes to the data shown in Tables 24 and 25 below.

Table 24: Wave 3 number of residents who smoke inside amended results – B cohort

B cohort	No. residents smoke inside - chb15a4a			
	Wave 3 (original)	%	Wave 3 (amended)	%
Refused (-3)	0	0.0	0	0.0
Not applicable (-9)	0	0.0	0	0.0
Missing (.)	1	0.0	12	0.3
0	4,060	92.6	4,174	95.2
1	136	3.1	136	3.1
2	58	1.3	58	1.3
3	3	0.1	3	0.1
4	1	0.0	1	0.0
5 or more	127	2.9	2	0.0
Total	**4,386**	**100.0**	**4,386**	**100.0**

Table 25: Wave 3 number of residents who smoke inside amended results – K cohort

K cohort	No. residents smoke inside - ehb15a4a			
	Wave 3 (original)	%	Wave 3 (amended)	%
Refused (-3)	0	0.0	0	0.0
Not applicable (-9)	0	0.0	0	0.0
Missing (.)	0	0.0	8	0.2
0	3,939	90.9	4,041	93.3
1	197	4.5	197	4.5
2	78	1.8	78	1.8
3	3	0.1	3	0.1
4	4	0.1	4	0.1
5 or more	110	2.5	0	0.0
Total	**4,331**	**100.0**	**4,331**	**100.0**

12 Missing data for Wave 6 items

12.1 Missing data for bullying items

In the Wave 6 ACASI instrument B cohort children were asked ACASB 6.1:

During the last 12 months, since [current month inserted] last year ...

a. kids hit or kicked me on purpose
b. kids grabbed or shoved me on purpose
c. kids threatened to hurt me
d. kids threatened to take my things
e. kids said mean things to me or called me names
f. kids tried to keep others from being my friend
g. kids did not let me join in what they were doing
h. kids used force to steal something from me
i. kids hurt me or tried to hurt me with a weapon
j. kids stole my things to be mean to me
k. kids forced me to do something I didn't want to do
l. I hit or kicked someone on purpose
m. I grabbed or shoved someone on purpose
n. I threatened to hurt someone
o. I threatened to take someone's things
p. I said mean things to someone or called someone names
q. I told others not to be someone's friend
r. I did not let someone join in what I was doing
s. I used force to steal something from someone
t. I hurt or tried to hurt someone with a weapon
u. I stole someone's things to be mean to them
v. I forced someone to do something they did not want to do

For selected categories children are then asked ACASB 6.5

How did this happen?

You can select ***more than one****. (If you are not using a mouse, press space bar between responses.)*

1. Face-to-face
2. Video chat
3. Phone call (not video chat)
4. Private messaging (includes email)
5. Open forum (e.g. Facebook walls, blogs, Twitter)
6. Other

Due to a sequencing error ACASB 6.5 was not asked for the following response categories of ACASB 6.1:

k. kids forced me to do something I didn't want to do

n. I threatened to hurt someone

o. I threatened to take someone's things

p. I said mean things to someone or called someone names

q. I told others not to be someone's friend

r. I did not let someone join in what I was doing

s. I used force to steal something from someone

v. I forced someone to do something they did not want to do

Thus, these items are not on the output file for the B cohort. The K cohort was not affected by this sequencing error.

12.2 Missing data for Cogstate items

In Wave 6, the K cohort study children were given tasks to assess executive functioning. Three tests were included:

- Identification task (testing visual attention and choice reaction time); for example, press Yes if card is Red, No if card is Black.
- One back task (testing working memory); for example, press Yes if current card is the same as the previous card, No if different.
- Groton Maze (spatial memory, impulse control and inhibition of erroneous responses); for example, trying to work out the correct path through a maze and remembering it for the next time through.

For some of the Wave 6 interviews, systems issues resulted in a small number of records with no executive functioning data on the output file. As Table 26 shows, the item hid40o1 indicates whether the executive functioning data are present and a reason for those records where the data are not present.

Table 26: Whether Cogstate data present – K cohort

14/15 - EXF- Cogstate data present - hid40o1	Number of records	%
Not applicable (-9)	146	4.1
Cogstate data present	3,234	91.4
Cogstate data not present – no consent given	43	1.2
Cogstate data not present – module could not be completed due to systems issues	83	2.4
Cogstate data not present – data loss due to systems issues	29	0.8
Cogstate data not present – child consented but did not complete any tasks	2	0.1

12.3 Missing data for puberty-related items

In Wave 6, all K cohort female study children were asked for the first time if they have ever menstruated (had your period). If they answered yes, then they were asked about age of first period, any periods in the last three months, and menstrual problems in the last three months. These questions were asked as part of the audio-computer-assisted self-interview (ACASI).

In previous waves, the P1 was asked whether the study child had ever menstruated and the age of first period. These questions were asked as part of the computer-assisted self-interview (CASI) mode of the interview.

In Wave 6, a roll forward error in the instrument meant not all eligible children (those who reported at Wave 6 that they had started their period) were asked the subsequent questions about age started period and period experiences in the last three months. Those cases where the P1 had previously reported that the study child's periods had started (n = 876) were not asked these additional questions.

Further, a sequencing error resulted in some study children (n = 85) being asked if they had a period in the last three months even though they answered 'no' to ever had a period.

The Wave 6 data were presented as collected from the study child. Items combining information collected from the two different informants have not been derived. Further, no data were amended when answers between the two informants conflicted or answers from the same informant conflicted. For example, when P1 reported that the study child's period had started, but the study child reported it hadn't, and when the study child reported they hadn't started period, but then reported one in last three months.

Table 27 presents the various items that could be used for analysis of this topic. Only K cohort items are presented given B cohort study children were not asked these questions in Wave 6.

Table 27: Study child menstruation items

Wave	Variable Name	Variable Label	Mode	Informant	Number missing
5	ghs36h	12/13 - CASI D4.7+W4 - Menstruate	CASI	Parent 1	NA
5	ghs36h1	12/13 - CASI D4.8+W4 - Menstruate (total months)	CASI	Parent 1	NA
6	hhs36h	14/15 - SC - ACASK 18.2 - Menstruate	ACASI	Study child	NA
6	hhs36h1	14/15 - SC - ACASK 18.3 - Menstruate (total months)	ACASI	Study child	876
6	hhs36i	14/15 - SC - ACASK 18.4 - Have you had any periods in the last 3 months	ACASI	Study child	876
6	hhs36i1	14/15 - SC - ACASK 18.5.1 - How regular were your periods	ACASI	Study child	876
6	hhs36i2	14/15 - SC - ACASK 18.5.2 - How heavy were your periods	ACASI	Study child	876
6	hhs36i3	14/15 - SC - ACASK 18.5.3 - How painful were your periods	ACASI	Study child	876
6	hhs36i4	14/15 - SC - ACASK 18.5.4 - How grumpy or teary did you get before your periods	ACASI	Study child	876
6	hhs36i5	14/15 - SC - ACASK 18.6.1 - Did you miss any school days	ACASI	Study child	876
6	hhs36i6	14/15 - SC - ACASK 18.6.2 - Did you miss any social activities	ACASI	Study child	876
6	hhs36i7	14/15 - SC - ACASK 18.6.3 - Did you miss any sports or exercise	ACASI	Study child	876

12.4 Missing data for study child helping others items

In Wave 6, K cohort study children were asked about helping others. More detailed questions about these caring situations were subsequently asked for up to three people. These questions were asked as part of the audio-computer-assisted self-interview (ACASI).

When study children entered more than three people when asked who they helped (e.g. Paul John George Ringo, instead of just Paul John George), the instrument was unable to process who to ask the subsequent questions about and as a result these questions were not asked (n = 132).

Table 28 (on page 53) lists the items affected and the counts. The second and third set of items were also set to missing, given it is unclear how many people the study children may have helped.

Table 28: Amount of missing data for study child helping others items

Variable Name	Variable Label	Number missing
hsc28a	14/15 - ACASK 12.1 - Help someone with long-term health condition/disability/elderly	NA
hsc28b1	14/15 - ACASK 12.3.1 - What is first person's relationship to you	132
hsc28b2	14/15 - ACASK 12.4.1 - Does first person live with you	132
hsc28b3	14/15 - ACASK 12.5.1 - Does first person go to the same school as you	132
hsc28e1	14/15 - ACASK 12.6.1.1 - Help provided - 1st person - Personal care	132
hsc28f1	14/15 - ACASK 12.6.1.2 - Help provided - 1st person - Moving around	132
hsc28g1	14/15 - ACASK 12.6.1.3 - Help provided - 1st person – Transport	132
hsc28h1	14/15 - ACASK 12.6.1.4 - Help provided - 1st person - Communicating	132
hsc28i1	14/15 - ACASK 12.6.1.5 - Help provided - 1st person - Preparation of meals	132
hsc28j1	14/15 - ACASK 12.6.1.6 - Help provided - 1st person - Housework/shopping/errands	132
hsc28k1	14/15 - ACASK 12.6.1.7 - Help provided - 1st person - House repairs/garden care	132
hsc28l1	14/15 - ACASK 12.6.1.8 - Help provided - 1st person - Health care	132
hsc28m1	14/15 - ACASK 12.6.1.9 - Help provided - 1st person – Paperwork	132
hsc28n1	14/15 - ACASK 12.6.1.10 - Help provided -1st person - Keeping them company	132
hsc28o1	14/15 - ACASK 12.6.1.11 - Help provided -1st person - Other	132
hsc28q	14/15 - ACASK 12.8 - How often do you do these caring activities	Not affected as not in a loop
hsc28r	14/15 - ACASK 12.9 - On average - total number of hours you spend providing care	Not affected as not in a loop

13 Issues with breadwinner questions

The Wave 7 instrument had an error in the information carried forward from Wave 5 and 6 that was used to sequence participants (P1 and P2) to the breadwinner questions, which were part of the parenting topic. As a result, only a very small percentage of those participants who were meant to be asked, were actually asked about, the non-breadwinner in their family (approximately 3%). Of this 3% there were also a small number of records with word substitution issues, which has meant there is some uncertainty if the data are about the right person (Mother or Father). As a result, a decision was made to drop the data from the main dataset and the following variables have been removed from the data dictionary:

Table 29: Variables removed from the data dictionary due to instrument error information carried forward from Wave 5 and 6 into Wave 7.

Variable name	Variable label
gpa24a2a	12/13 - P1 - F2F B8.6.1+W4 - Parent 1's father worked
gpa24a5a	12/13 - P1 - F2F B8.7.1+W4 - Parent 1's father's occupation
gpa24a4a	12/13 - P1 - F2F B8.8.1.1+W4 - Parent 1's father unemployed for 6 months
gpa24a2b	12/13 - P1 - F2F B8.6.2+W4 - Parent 1's mother worked
gpa24a5b	12/13 - P1 - F2F B8.7.2+W4 - Parent 1's mother's occupation
gpa24a4b	12/13 - P1 - F2F B8.8.1.2+W4 - Parent 1's mother unemployed for 6 months
ipa24a2a	16/17 - P1 - P1 CAI B10.1.2+W4-6 - P1's father work in a job, business or farm?
ipa24a5a	16/17 - P1 - P1 CAI B10.1.3/10.1.4+W4-6 - P1's father's occupation
ipa24a4a	16/17 - P1 - P1 CAI B10.1.5+W4-6 - P1's father unemployed for 6 months or more
ipa24a2b	16/17 - P1 - P1 CAI B10.1.2+W4-6 - P1's mother work in a job, business or farm?
ipa24a5b	16/17 - P1 - P1 CAI B10.1.3/10.1.4+W4-6 - P1's mother's occupation
ipa24a4b	16/17 - P1 - P1 CAI B10.1.5+W4-6 - P1's mother unemployed for 6 months or more
gpa24b2a	12/13 - P2 - F2F B8.6.1+W4 - Parent 2's father worked
gpa24b5a	12/13 - P2 - F2F B8.7.1+W4 - Parent 2's father's occupation
gpa24b4a	12/13 - P2 - F2F B8.8.2.1+W4 - Parent 2's father unemployed for 6 months
gpa24b2b	12/13 - P2 - F2F B8.6.2+W4 - Parent 2's mother worked
gpa24b5b	12/13 - P2 - F2F B8.7.2+W4 - Parent 2's mother's occupation
gpa24b4b	12/13 - P2 - F2F B8.8.2.2+W4 - Parent 2's mother unemployed for 6 months
ipa24b2a	16/17 - P2 - P1 CAI B10.2.2+W4-6 - P2's father work in a job, business or farm?
ipa24b5a	16/17 - P2 - P1 CAI B10.2.3/10.2.4+W4-6 - P2's father's occupation
ipa24b4a	16/17 - P2 - P1 CAI B10.2.5+W4-6 - P2's father unemployed for 6 months or more
ipa24b2b	16/17 - P2 - P1 CAI B10.2.2+W4-6 - P2's mother work in a job, business or farm?
ipa24b5b	16/17 - P2 - P1 CAI B10.2.3/10.2.4+W4-6 - P2's mother's occupation
ipa24b4b	16/17 - P2 - P1 CAI B10.2.5+W4-6 - P2's mother unemployed for 6 months or more
gpa24m2a	12/13 - M - F2F B8.6.1+W4 - Mother's father worked

Table continued on next page →

Variable name	Variable label
gpa24m5a	12/13 - M -F2F B8.7.1+W4 - Mother's father's occupation
gpa24m4a	12/13 - M - F2F B8.8.1+W4 - Mother's father unemployed for 6 months
gpa24m2b	12/13 - M - F2F B8.6.2+W4 – Mother's mother worked
gpa24m5b	12/13 - M - F2F B8.7.2+W4 - Mother's mother's occupation
gpa24m4b	12/13 - M - F2F B8.8.2+W4 - Mother's, mother unemployed for 6 months
ipa24m2a	16/17 - M - P1 CAI B10.2.2+W4-6 - Mother's father work in a job, business or farm
ipa24m5a	16/17 - M - P1 CAI B10.2.3/10.2.4+W4-6 - Mother's father's occupation?
ipa24m4a	16/17 - M - P1 CAI B10.1.5/10.2.5+W4-6 - Mother's father unemployed for 6 months or more
ipa24m2b	16/17 - M - P1 CAI B10.2.2+W4-6 - Mother's mother work in a job, business or farm?
ipa24m5b	16/17 - M - P1 CAI B10.2.3/10.2.4+W4-6 - Mother's mother occupation?
ipa24m4b	16/17 - M - P1 CAI B10.1.5/10.2.5+W4-6 - Mother's mother unemployed for 6 months or more
gpa24f2a	12/13 - F - F2F B8.6.1+W4 - Father's father worked
gpa24f5a	12/13 - F - F2F B8.7.1+W4 - Father's father occupation
gpa24f4a	12/13 - F - F2F B8.8.1+W4 - Father's father unemployed for 6 months
gpa24f2b	12/13 - F - F2F B8.6.2+W4 - Father's mother worked
gpa24f5b	12/13 - F - F2F B8.7.2+W4 - Father's mother's occupation
gpa24f4b	12/13 - F - F2F B8.8.2+W4 - Father's mother unemployed for 6 months
ipa24f2a	16/17 - F - P1 CAI B10.2.2+W4-6 - Father's father work in a job, business or farm?
pa24f5a	16/17 - F - P1 CAI B10.2.3/10.2.4+W4-6 - Father's father's occupation?
ipa24f4a	16/17 - F - P1 CAI B10.1.5/10.2.5+W4-6 - Father's father unemployed for 6 months or more
ipa24f2b	16/17 - F - P1 CAI B10.2.2+W4-6 - Father's mother work in a job, business or farm?
ipa24f5b	16/17 - F - P1 CAI B10.2.3/10.2.4+W4-6 - Father's mother's occupation?
ipa24f4b	16/17 - F - P1 CAI B10.1.5/10.2.5+W4-6 - Father's mother unemployed for 6 months or more

Notes: 12/13 = Wave 7 B cohort. 16/17 = Wave 7 K cohort.

There was also an issue with the information carried forward into the Wave 7 PLE CATI that was used to sequence participants to the questions asking about the non-breadwinner in their family. Due to this error, **no** participants were sequenced to these questions. As a result, there are no data for these questions and the following variables have been removed from the data dictionary:

Table 30: Variables removed from the data dictionary due to data issue with PLE CATI information carried forward into Wave 7

Variable name	Variable label
ipa24p2b	16/17 - PLE 30.1+W4-6 - PLE's mother work in a job, business or farm when PLE 14 years old
ipa24p5b	16/17 - PLE 30.2/30.3+W4-6 - PLE's mother's occupation
ipa24p4b	16/17 - PLE 30.4+W4-6 - PLE's mother unemployed for 6 months or more when PLE 14 years old
ipa24p2a	16/17 - PLE 30.1+W4-6 - PLE's father work in a job, business or farm when PLE 14 years old
ipa24p5a	16/17 - PLE 30.2/30.3+W4-6 - PLE's father's occupation
ipa24p4a	16/17 - PLE 30.4+W4-6 - PLE's father unemployed for 6 months or more when PLE 14 years old
gpa24p2a	12/13 - PLE 30.1+W4-6 - PLE's father work in a job, business or farm when PLE 14 years old
gpa24p5b	12/13 - PLE 30.2/30.3+W4-6 - PLE's mother's occupation
gpa24p4b	12/13 - PLE 30.4+W4-6 - PLE's mother unemployed for 6 months or more when PLE 14 years old
gpa24p2b	12/13 - PLE 30.1+W4-6 -PLE's mother work in a job, business or farm when PLE 14 years old
gpa24p5a	12/13 - PLE 30.2/30.3+W4-6 - PLE's father's occupation
gpa24p4a	12/13 - PLE 30.4+W4-6 - PLE's father unemployed for 6 months or more when PLE 14 years old

Notes: 12/13 = Wave 7 B cohort. 16/17 = Wave 7 K cohort.

14 Date of birth corrections

At each interview, or between interviews, parents may advise us of a correction to the date of birth of anyone in the household (usually the study child). These are reflected in the items zf04mx, where x is the member number. So zf04m1 would be the variable for the SC for this example.

In addition to these updates, staff from the Murdoch Children's Research Institute noticed some date of birth discrepancies (compared to LSAC recorded date of birth) while conducting the Child Health CheckPoint (CHCP) sessions.

Due to the quite robust nature of the checkpoint process for obtaining and checking date of birth, the ABS has investigated these discrepancies (including calling the participant in some cases) and updated the LSAC date of birth in the Wave 7 Household file where appropriate.

There has been a total of 82 changes to study child date of birth since Wave 1. To assist users in identifying these records, an additional indicator variable has been added to the Household file and the data dictionary. It has been linked with the household member number to maximise its usefulness. The variable is zf04am*. Users can enter desired member number at * to return the data required (e.g. M1 for SC, M2 for P1 and M3 for P2).

Table 31 shows the magnitude for those changes and if the change came during Waves 1–6 (or from the CHCP data checks).

Table 31: Shows changes to study child date of birth during Wave 1 to 6 (or from the CHCP data check)

Date of birth change	Number of SC records corrected during Waves 1–6	Wave 7 from CHCP	Total
Less than 1 month	43	4	47
1 month < 3 months	15	2	17
3 months < 6 months	10	0	10
6 months < 12 months	5	0	5
12 months or more	3		3
Total	**76**	**6**	**82**

Wave 7 corrections have been made and any derived variables affected have been calculated using the correct date of birth. Wave 1–6 updates have been made to the datasets so that any derived variables affected by a change to date of birth have now been re-derived using the correct date of birth. The only exception to this is the Academic Rating Scale (ARS) data items. These have not been updated for Waves 1–6 due to a cost–benefit decision. Users of these data can make use of the new date of birth change indicator variable to exclude records if necessary.

15 Minor changes for weight, BMI and height percentiles and z-scores

As part of the investigation into the date of birth corrections (outlined in section 14), some inconsistency was discovered in the method for calculating 'SC age in months' across the waves:

- Wave 1 and Wave 2 were calculated using the integer method
- Wave 3 and Wave 4 were calculated using a custom method that correctly derived the number of completed months to determine age in months
- Wave 5 and Wave 6 were calculated using the rounding method.

The calculation method for Waves 1, 2, 5 and 6 has now been changed to match the custom method used for Waves 3 and 4. This method was chosen for all waves as it correctly derives the number of 'true' completed months (using the date of interview as a reference point for age calculation) as opposed to the integer method that simply divides the number of days since the interview by 30.44.

As result of this method change, 'age in months for SC' was recalculated for Waves 1, 2, 5 and 6. A large number of records were altered by +1 or –1 for age in months in these waves. However, only very minor changes were observed, when before and after distributions were compared.

16 Body fat percentage data corrections

There was a data processing error in the treatment of body fat values in Wave 5 and Wave 6 for both cohorts. The variables affected were:

- Wave 5 B cohort - ebodyfat
- Wave 5 K cohort - gbodyfat
- Wave 6 B cohort - fbodyfat
- Wave 6 K cohort – hbodyfat

This error was due to the incorrect treatment of the decimal place for records where the interviewer did not enter the zero after the decimal point. For example, if the interviewer entered a value of 27, it resulted in a processed value in the final data of 2.7 rather than 27.0.

All records affected by this incorrect treatment have been corrected for the Wave 7 data release. The number of records corrected are shown in Table 32.

Table 32: The number of records corrected for B and K cohorts by wave

Cohort	Wave 5	Wave 6
B cohort	171	215
K cohort	164	156

17 Wave 4 salary and wages

There was an issue in the Wave 4 data instrument for wages and salaries questions. If a participant's only source of income identified (at the income source questions fn02a) was wages and salaries, then the instrument did not ask the participant the subsequent wages and salary questions (fn13a). It used the usual income amount (fn09) as the wages and salaries amount.

The Wave 4 data release was not showing the salary data and salary groups for these participants who had wages and salaries as their only source of income. This has been corrected for the most recent release.

Also, those with 'no income' were missing from the income group items. This has also been corrected so that they now are included in the group.

18 Study children allergies (issues with Wave 6 and 7 data)

In Waves 5, 6 and 7 the P1s in B and K cohorts were asked about the study child's existing allergies to food items and also asked if there had been any reactions to food items since the previous interview. The relevant questions are HEAL 17 and 18 with the word sub 'allergy' sourced from roll forward (pre-fill) data:

- HEAL_Q17* (hs43*) – In a previous interview, you told us that [(SC)] had a reaction to [(allergy)].
- HEAL_Q17* (hs43*) – Has [(SC)] had any more reaction to [(allergy)] since we last saw you?
- HEAL_Q18* (hs39*) – Since the last interview, has [SC] had a reaction (e.g. redness or itching) that you thought was due to some food or drink [SC] had consumed?

While processing the Wave 7 data some issues were discovered that originated in errors made in the roll forward (pre-fill) process. These errors have affected the Waves 6 and 7 data for some records in B and K cohorts. For some records, the historical allergy information was not rolled forward correctly resulting in the participant not being asked when they should have been. There were also output errors in the release data for Wave 6. The following section contains detailed explanations of the issues.

18.1 Wave 6 data issues

Roll forward from Wave 5 to 6 (Pre-existing allergies HEAL 17)

For Other Nuts, Milk, Wheat and Other Allergies there were data roll forward errors. If a respondent answered 2 (Withdrawn food) or 3 (Still have allergy) in Wave 5, the response should have been rolled forward to pre-fill the Wave 6 instrument. This wasn't done due to processing errors that occurred while producing the pre-fill files. This meant that for these allergies Q17* was not asked and therefore data on whether the study child continued to experience these allergies is absent from Wave 6. The number of records affected are shown in Table 33.

Table 33: Wave 6 records affected by pre-fill errors

	Other nuts	Milk	Wheat	Other allergies
B cohort (records affected)	28	39	14	170
K cohort (records affected)	27	25	13	158

For all other allergy types (Peanut, Eggs, Soy, Sesame) the roll forward of the pre-fill information was correctly done.

Roll forward from Wave 5 to 6 for new reactions (HEAL 18)

All correct.

Error in output data for Wave 6 – HEAL 17

There was an error in Wave 6 output data for the responses to questions about existing allergies (HEAL 17). The addition of a number of other allergy categories or collection caused a mapping error from the input data through to the output data. The input data is correct for all allergies (not affected by the roll forward data problem) but the input data for the following allergy types were mapped to the incorrect output item:

- Fruit
- Preservative
- Additive
- Seafood
- Other allergies.

The number of records affected are shown in Table 34.

Table 34: Records affected by output mapping errors

	Fruit	Preservative	Additive	Seafood	Other allergies	Total
B cohort (records affected)	24	27	34	9	13	107
K cohort (records affected)	15	10	16	11	19	71

These have been corrected for the Wave 6 data as part of the Wave 7 data release.

A flow-on from this error was that the incorrect version of the output data was used as the basis of the pre-fill information for the Wave 7 allergy information, so it has also had an effect on the quality of the Wave 7 data (see Wave 7 data section below).

18.2 Wave 7 data issues

Roll forward from Wave 6 to 7 (Pre-existing allergies HEAL 17)

The creation of the roll forward file for Wave 7 relies on the use of the output data to ensure the most current data are used. As mentioned above, the output file for Wave 6 was incorrect (due to the output mapping issue). So, all of the roll forward indictors for Fruit, Preservative, Additive, Seafood and Other allergies were incorrect – 107 records in B cohort and 71 records in K cohort were affected as a result of no indicator data being rolled forward for any of these allergy types (as per Table 34 above.)

There was also another separate processing issue with the roll forward of Milk where some participants in Wave 7 were asked about existing milk allergies incorrectly; that is, the study child did not have an existing allergy to milk but was asked Q17* about whether they had experienced any further reactions in error at interview (24 records in B cohort and 6 records in K cohort). This error occurred due to a processing data extraction issue, not as a result of the incorrect mapping problem described above.

For all other allergy types (Peanut, Other nuts, Eggs, Soy, Sesame and Wheat) the roll forward was correctly done between Waves 6 and 7. For records where roll forward was done incorrectly between Wave 5 and Wave 6 for Other Nuts, Milk, Wheat and Other allergies, the error has carried through to Wave 7 data; for example, those records were still not asked about their existing allergy from Wave 5.

Roll forward from Wave 6 to 7 for new reactions (HEAL 18)

All correct.

18.3 Corrections made

In relation to the mapping issue, Wave 6 output data have been fixed so that the input data are correctly reflected in the output items in the released data files.

In relation to the Wave 5 to Wave 6 roll forward issue, any data items for records that were in scope to be asked about known pre-existing allergies (including the subsequent reaction type variables), but were not asked, were set to missing.

Any data items (including the subsequent reaction type variables) for records that were asked about an existing milk allergy in error were set to -9 (not asked).

19 After school care issue Wave 7 B cohort

19.1 Variable name pc02e and pc02eo

The variable name pc02e and pc02eo refers to 'main reason child does not have any regular child care arrangements before and or after school'.

From Wave 4 through to Wave 7, the data regarding why the study child does not have regular child care arrangements was asked as follows:

- Wave 4: question was asked about current wave.
- Wave 5: asked as a catch-up 'two years ago' (i.e. asking about Wave 4 period). Data was combined for data release (W4 + W5) and used previous Wave age indicator and pre-Wave age in SC age column.
- Wave 6: asked as a catch-up question again 'two years ago' (i.e. asking about Wave 5 period) but there are no Wave 5 'current' data to be combined with (only catch-up data for a different period).
- Wave 7: asked as a catch-up question again 'two years ago' (i.e. asking about Wave 6 period) but there is no Wave 6 'current' data to be combined with (only catch-up data for a different period).

Based on the history of the questions, Waves 6 and 7 are not 'correct' catch-up questions as they don't relate to the Wave 4 time period and only ever asked participants who missed answering in previous waves. These items have become unusable in Wave 6 and Wave 7 due to the reducing population covered each time. As a result, the items have been dropped from the Wave 7 datasets and data dictionary.

20 Who is mother/father issue

Since Wave 4, the study child has completed their own computer-assisted interview on a separate notebook from their parents. They have been asked questions about their relationships with their parents and their opinions on parents' work.

At the start of the relevant module, interviewers enter information about the family to determine who the study child will report on in relation to their mum/dad, taking into account the fact that parents could be living elsewhere, deceased or multiple people of the same sex may be playing a parental role.

Where a study child has their biological/adopted mother and father in the home, the words 'your mum/your dad' are used for word substitutions in follow-up questions. However, if the biological/adopted mother and father is living elsewhere or deceased/never sees, the study child is asked to enter the name of the person that the follow-up questions will relate to. In initial data releases, users were not made aware of which person was listed in these follow-up questions.

In addition, there were some minor instrument issues in Wave 6 and Wave 7. In Wave 6 the explicit interviewer question as to whether the study child chose their 'mum/dad living elsewhere' was not asked at all (question wasn't included in Wave 5). In Wave 7 sequencing issues meant that the SCCASI MumSkip question was not asked at all, which means that the majority of records have no specific answers as to who the study child is referring to in regards to 'Mum' data.

Table 35 provides a summary of the introduction questions to the relevant module for each cohort.

Table 35: Biological/adopted mother/father in the home according to ACASI/CSR introduction questions for each cohort

	Wave 5		Wave 6		Wave 7	
	Mother	Father	Mother	Father	Mother	Father
B cohort						
Yes, in this home	**3,963**	**3,340**	**3,542**	**2,913**	**3,144**	**2,559**
No, living elsewhere (PLE)						
More than one person could be a mum/dad = Yes (a)	7	135	17	143	16	142
More than one person could be a mum/dad = No	*38*	*488*	*48*	*506*	*38*	*451*
Total	*45*	*623*	*65*	*649*	*54*	*593*
No, deceased or never sees						
Is there a mum/dad in the home = Yes (b)	3	8	6	13	8	12
Is there a mum/dad in the home = No	5	45	3	41	7	49
Total	*8*	*53*	*9*	*54*	*15*	*61*
Total	**4,016**	**4,016**	**3,616**	**3,616**	**3,213**	**3,213**
Total needing follow-up question about name (a + b)	**10**	**143**	**23**	**156**	**24**	**154**
K cohort						
Yes, in this home	**3,731**	**3,056**	**3,224**	**2,627**	**2,804**	**2,269**
No, living elsewhere (PLE)						
More than one person could be a mum/dad = Yes (a)	22	191	16	167	NA*	NA*

Table continued on next page →

	Wave 5		Wave 6		Wave 7	
	Mother	Father	Mother	Father	Mother	Father
More than one person could be a mum/dad = No	79	520	83	459	NA*	NA*
Total	*101*	*711*	*99*	*626*	*91*	*563*
No, deceased or never sees						
Is there a mum/dad in the home = Yes (b)	7	11	8	15	2	10
Is there a mum/dad in the home = No	14	75	18	81	18	73
Total	*21*	*86*	*26*	*96*	*20*	*83*
Total	**3,853**	**3,853**	**3,349**	**3,349**	**2,915**	**2,915**
Total needing follow-up question about name (a + b)	**29**	**202**	**24**	**182**	**2****	**10****

Notes: NA* - These values are unavailable due to question not being asked for Wave 7. ** These values have been affected due to the unavailability of the values at Row 3 and 4 of the K cohort in the table.

Tables 36 and 37 provide the results of the follow-up matching process undertaken on the names entered by the study child in the introduction questions for each cohort.

Table 36: Results for matched names entered by the study child in the introduction questions for B cohort

B cohort	Wave 5		Wave 6		Wave 7	
	Mother	Father	Mother	Father	Mother	Father
Matched						
Current Wave PLE	3	79	2	79	3	86
Current Wave P2	2	42	8	50	6	48
Current Wave P1	3	0	5	0	9	0
Other HH member	0	4	1	4	0	0
Total	**8**	**125**	**16**	**133**	**18**	**134**
Not matched						
No match to names on HHF	1	2	0	1	3	3
Not specific	1	9	1	9	3	8
Nothing entered	0	7	6	13	0	9
Total	**2**	**18**	**7**	**23**	**6**	**20**
Total	**10**	**143**	**23**	**156**	**24**	**154**

Table 37: Results for matched names entered by the study child in the introduction questions for K cohort

K cohort	Wave 5		Wave 6		Wave 7	
	Mother	Father	Mother	Father	Mother	Father
Matched						
Current Wave PLE	10	101	6	68	0	0
Current Wave P2	4	58	6	57	1	10
Current Wave P1	7	1	3	1	0	0
Other HH member	1	5	0	5	0	0
Total	**22**	**165**	**15**	**131**	**1**	**10**
Not matched						
No match to names on HHF	0	7	1	3	0	0
Not specific	2	15	2	15	1	0
Nothing entered	5	15	6	33	0	0
Total	**7**	**37**	**9**	**51**	**1**	**0**
Total	**29**	**202**	**24**	**182**	**2**	**10**

Note: *These numbers are a large undercount due to instrument issues in Wave 7 as outlined earlier.

The datasets have been amended to either include a new item or update an existing item. Table 38 shows which values have been given in each case.

Table 38: Values that have been given in each case

	fd23d1/fd23d1* (Study child chose their mum/dad living elsewhere)	fd23e1/fd23e1* (Who study child chose to respond about for mother/ father)
Matched		
Current Wave PLE	Yes	Current Wave PLE
Current Wave P2	No	Current Wave P2
Current Wave P1	No	Current Wave P1
Other HH member	No	Other/unknown
Not matched		
No match to names on HHF	Don't Know	Other/unknown
Not specific	Don't Know	Other/unknown
Nothing entered	Don't Know	Other/unknown

Note: *Wave identifier not included, either e/f/g/h.

21 Repeated a year level issue

There was an instrument issue in the B cohort Education module for Wave 7. An error in the sequencing specifications meant that the records that had a difference of one year in their 2014 and 2016 grade levels were not asked the questions regarding any repeats of grade/level since last interview. These participants with a difference of only one year in their 2014 and 2016 grade levels were also the most likely to respond in the positive to the repeat year questions.

As a result, the data for which grade/year level was repeated (gpc47a2) and what was the main reason, or other reason, for repeating the grade/level (gpc47a3a, gpc47a3b) have been dropped from the Wave 7 data file. The data for whether a grade has been repeated or not (gpc47a6) has been derived for B cohort by using the grade indicated by the study child in Wave 6 and Wave 7.

22 Executive functioning – CogState – missing data Wave 7

The executive functioning of children in the K cohort was tested at Wave 6 using three Cogstate cognitive tests, including the Identification task (IDNT), One-back test (ONBT), and Groton Maze Learning Test (GML). In Wave 7, the same battery of tests was used to examine the executive functioning of the P1 of K cohort children. The outcome variables are contained in the CogState dataset, where a series of cognitive testing batteries have been customised for use in LSAC. Each row of a CogState dataset represents one task in the CogState test battery for one study subject in one test session. Further information is available in the *Data User Guide, Wave 8* (growingupinaustralia.gov.au/data-and-documentation/data-user-guide), and the LSAC Technical Paper No. 19, *Executive Functioning – Use of Cogstate measures in the Longitudinal Study of Australian Children* (available from the study website growingupinaustralia.gov.au/data-and-documentation/technical-papers).

At Wave 7, 465 parents (15% of Wave 7 responding sample) had missing CogState data. Reasons for parents' CogState data being missing are presented in Table 39.

Table 39: CogState interviews for the K cohort in Wave 7

	Number	Percentage
Not in scope		
SC only interviews (RAP) *	13	0.4
SC only interviews (Non - RAP) *	15	0.5
7.25 interviews *	13	0.4
In scope		
1. CogState data present	2,624	84.9
2. CogState data not present – no consent given *	312	10.1
3. CogState data not present – module could not be completed due to systems issues *	32	1.0
4. CogState data not present – data loss due to systems issues *	61	2.0
5. CogState data not present – consented but did not complete any tasks *	10	0.3
6. CogState data not present – partially responding record, didn't get to exec. module *	9	0.3
Total interviews for K cohort in Wave 7	**3,089**	**100.0**

Note: *Sum of 465 interviews.

The most common reason for missing CogState data was no P1 consent. The level of consent was lower among parents (89.1% of responding Wave 7 sample) than study children (99.5% of responding Wave 6 sample).

Table 40: CogState data not present – breakdown of reasons for consent not given

Reasons	Number	Percentage
Time constraints	99	31.7
Don't want to do a test/puzzle	83	26.6
Computer literacy	24	7.7
Tired	23	7.4
No reason	20	6.4
Illness/injury	19	6.1
Telephone interview	18	5.8
Environment Issues	12	3.8
Language	8	2.6
Technical difficulties	6	1.9
Total	**312**	**100.0**

During Wave 7 enumeration, executive functioning files were obtained and loaded monthly to speed up processing and also to ensure the software and ABS systems were working properly. Through this early processing, the ABS discovered two issues accounting for records in categories 3 and 4 in Table 39.

22.1 Category 3: CogState data not present – module could not be completed due to systems issues

The module could not be completed for two main reasons: (1) one interviewer had not had CogState software installed on the laptop (accounting for 11 missing records); and (2) the software kept crashing at the time of testing (accounting for the rest of records).

22.2 Category 4: CogState data not present – data loss due to systems issues

Some records had missing executive functioning files (encrypted files containing the executive functioning data) even though it appeared that the CogState software had run successfully. The common pattern for affected records was a combination of a male Parent 1 and a female study child. In these circumstances, incorrect instrument coding caused the sex of the person completing the tasks to be missing. This resulted in the executive functioning file not being created.

Once this issue was uncovered during the enumeration process, interviewers were instructed to change the sex of P1 to female for any remaining records that had a combination of male P1 and female SC. This enabled the executive functioning module to work and processing staff subsequently changed the sex of P1 back to male. There were 78 Wave 7 interviews that were affected by this issue. Of these:

- Thirteen did not consent to complete the executive functioning tasks.
- Thirteen successfully had the sex of P1 changed by the interviewer to female to allow the executive functioning module to work.
- Fifty records did not have the sex changed and therefore there is no executive functioning data.
- Two did not work for other reasons (category 3, system crash or software missing).

In addition, there were 11 records where the P1 sex was changed from female to male due to a role change in the household or because the sex was incorrect. This also resulted in data loss.

23 Expected/received child support per child

From Wave 1 through to Wave 4, the items 'Monthly child support expected' (pe20a1, pe20p1) and 'Child support received last month' (pe20a2, pe20p2) were based on child support for a child per period. The amount for a child was calculated by dividing the amount received or paid, for that period, by the number of children that received support.

However, in Wave 5 to Wave 6 these items were not calculated per child but rather collected as the total amount received, or paid, for that period.

The amount each child receives can be different if they are from different biological parents so the per-child values in Waves 1–4 are an average value only.

These differences are relevant to both P1 and PLE items in Wave 5 and Wave 6.

The per child values in Waves 1–4 will be left as they are but users should note the possible inaccuracies in the amounts if children have different biological parents. For Wave 5, and all subsequent waves, only the total amount received or paid for that period will be provided. A note has also been placed in the data dictionary to assist users in creating their own calculation of 'per child items' for Wave 5 onwards.

24 Reason for change in education institution – SC CAI 6.5

In Wave 7 in the CAI instrument for the K cohort, participants were asked:

'What is the main reason for the most recent change of education institution?'

This question was limited to the following population:

- Study child was away from home (RAP);
- Study child was studying; and
- Study child had changed schools.

Only a very small number of the Wave 7 K sample would have been sequenced to the question but analysis after enumeration showed that no responding participants had answered this question. As a result, this item (pc44c3b1) has been dropped from the data dictionary and data files for Wave 7.

25 Child support – parent living elsewhere PLE 20.8

In Wave 7 in the PLE CATI the parent living elsewhere should have been asked about looking after the study child when needed. Due to an error in sequencing, participants were not asked this question. As a result, this item (pe21p5) has been dropped from the data dictionary and data files for Wave 7.

26 Informant indicator in LSAC variable naming convention: Approach in Wave 7 and subsequent Waves

LSAC study content was asked almost exclusively from parents in the early waves (prior to Wave 5). As the LSAC study children are getting older, from Wave 7, some of the questions that were previously asked of their parents are now asked of the young people themselves. Some of the variables that were previously reported by parents in Wave 1 to 6 do not contain an informant indicator. The informant/subject indicator in the variable name is essential to LSAC because if more than one informant is reporting on the same questions, then the variable names will be different only by informant indicator (6th digit in the variable name). Therefore, variable names with informant indicators allow differentiating between data received for different respondents.

LSAC data management explored various possibilities to bridge existing gaps including the introduction of new variable names for parent and child variables with the informant indicators for Wave 7 (i.e. 'a' = P1, 'b' = P2, 'c'=SC etc.) and rename parent reported variables in Waves 1 to 6.

LSAC critically examined the benefits and consequences of the renaming variable approach for data users. The ongoing nature of renaming variables in subsequent waves challenges the longitudinal consistency of parent-reported variables across waves and also restricts the utilisation of data users' existing syntax files to reproduce analyses and outcomes for later waves.

Rather than renaming the existing variables that do not have an informant indicator, LSAC will continue to use the existing variable names for parent-reported variables (for previous and future waves); and follow a naming convention with an informant indicator for the new Study Child reported variables from Wave 7 onward. This new approach will ensure the longitudinal consistency of parent-reported variables without jeopardising the reproducibility of data users' existing code.

27 Desired occupation sequencing issue

As part of the SC CAI Wave 7 SC K cohort participants were asked 'Since the (last interview/In the last two years) have there been any times when you were actively looking for work?' (ipw41c1).

Respondents who answered 'no' at this point were sequenced incorrectly and were not asked the desired occupation question: 'As things stand now, do you know what career or occupation you would like to have in the future?' (ipw39ca).

Respondents who answered in the positive to actively looking for work were then asked if they had been actively looking for full-time or part-time work during the last four weeks (ipw11c4). Respondents who answered 'no' to this question were also incorrectly sequenced away from the desired occupation question.

As a result, only 509 out of a possible 3,089 were actually asked the desired occupation question. For Wave 6 a total of 3,317 respondents were asked what their desired occupation was. The decision was made to still output the data for the restricted population and place a note in the data dictionary to alert users to the smaller population for this question for Wave 7.

28 Inconsistent placement of SC question

The following question, about who study children talk to about their plans for the future, has been included for the K cohort in Waves 6, 7 and 8:

When you talk about your plans for the future, would you say you talk to your ...

10. Parents
11. Brother/sister*
12. Other relative/family member
13. School Career Guidance Counsellor
14. Psychologist/therapist*
15. Coaches/instructors*
16. Teachers
17. People from work (e.g. colleagues, employer)*
18. Boyfriend/Girlfriend/Partner*
19. Friends
20. Other unrelated adults (e.g. family friends, friend's parents)*
21. Do not talk to anyone*
22. Nobody to talk to*
23. Other

Note: *Responses added at Wave 7

The wording of the question has been the same across the three waves, with additional response categories added for Wave 7. However, the position of the question has changed between waves and this is likely to have implications for longitudinal comparability. In Wave 6 this question was placed after questions related to education, therefore the question is likely to have been answered with future education plans in mind. In Wave 7 the question was positioned after questions on desired future occupation, most likely resulting in responses being about work-related future plans. Due to the inconsistencies in placement, cross-Wave comparisons of these items need to be treated with caution as any changes found may be due to changes in the type of future plans being considered rather than actual shifts in who study children talk to.

29 Difference in health status of household members across waves of LSAC

It is important to note that questions about restrictive health conditions taken from the LSAC Household Form, which asked about every member of the study child's household, are not consistent across waves. Therefore, it cannot be used to examine changes in the health status and restrictive health conditions of household members over time. For example, in Wave 1, the study child's primary carer was asked: '*Does [person] have any medical conditions or disabilities that have lasted, or are likely to last, for six months or more?*'. The interviewer note stated: 'This refers to physical, psychological and emotional conditions that have lasted or are likely to last for six months or more. Often these conditions need medical treatment or assistance and affect everyday life.' However, in Wave 1, respondents were not asked if these conditions caused any restriction in the household members' everyday activities.

The prompt card provided with this question listed the following conditions:

- Sight problems (not corrected by glasses or contact lenses)
- Hearing problems
- Speech problems
- Blackouts, fits or loss of consciousness
- Difficulty learning or understanding things
- Limited use of arms or fingers
- Difficulty gripping things
- Limited use of legs or feet
- Nervous or emotional conditions that require treatment
- Any disfigurement or deformity
- Chronic or recurring pain
- Any condition that restricts physical activity or physical work (e.g. back problems, migraines)
- Shortness of breath or difficulty breathing
- Any mental illness for which help or supervision is required
- Long-term effects as a result of head injury, stroke or other brain damage
- Any other long-term condition such as arthritis, asthma, heart disease, dementia, etc.
- Any other long-term condition that requires treatment or medication

From Wave 2 onwards, these questions are separated into two components, the first asking about health conditions or disabilities that have lasted for 12 months and the second asking about conditions that restrict everyday activities.[5] The following two questions were asked in relation to each household member:

5 Across all waves (including Wave 1), the naming convention and variable labels for these variables is the same. Variables #f17am* to #f17jm* represent medical conditions or disabilities from sight problems/loss of sight to disfigurement or deformity. Variables #f18am* to #f18gm* represent conditions that may restrict everyday activities – ranging from "difficulty breathing" to "other treated condition". (Where # is a Wave indicator and * represents household member number).

1. *Does [person] have any medical conditions or disabilities that have lasted, or are likely to last, for six months or more?*

In Waves 2, 5, 6 and 7, the conditions listed on the prompt card for this question were:

Sight problems (not corrected by glasses or contact lenses)

Hearing problems (where communication is restricted, or an aid to assist with or substitute for hearing is used)

Speech problems

Blackouts, fits or loss of consciousness

Difficulty learning or understanding things

Limited use of arms or fingers

Difficulty gripping things

Limited use of legs or feet

Any condition that restricts physical activity or physical work

Any disfigurement or deformity.

In Waves 3 and 4, the conditions listed on the prompt card for this question were:

Loss of sight (not corrected by glasses or contact lenses)

Loss of hearing (where communication is restricted, or an aid to assist with or substitute for hearing is used)

Speech difficulties

Blackouts, fits or loss of consciousness

Difficulty learning or understanding things

Incomplete use of arms or fingers

Difficulty gripping or holding things

Incomplete use of legs or feet

Restriction in physical activities or doing physical work

Disfigurement or deformity.

2. *Still thinking of conditions lasting six months or more, is [person] restricted in everyday activities because of any of the following?*

In Waves 2, 5, 6 and 7, the conditions listed on the prompt card for this question were:

Shortness of breath or breathing difficulty

Chronic or recurring pain

A nervous or emotional condition (requiring treatment)

Any mental illness for which help or supervision is required **long-term**

Long-term effects as a result of a head injury, stroke or other brain damage

Any **other long-term condition, such as arthritis, asthma, heart disease, Alzheimer's, dementia**, etc.

Any other long-term disease or condition **that requires treatment or medication**.

In Waves 3 and 4, the conditions listed on the prompt card for this question were:

Shortness of breath or breathing difficulties **causing restriction**

Chronic or recurring pain or discomfort **causing restriction**

A nervous or emotional condition **causing restriction**

Mental illness or condition requiring help or supervision

Long-term effects of head injury, stroke or other brain damage **causing restriction**

Receiving treatment or medication for any long-term conditions or ailments and still restricted

Any other long-term conditions **resulting in a restriction**.

The separation of this question into two parts from Wave 2 onwards, and the differences in the wording on the prompt cards between waves have resulted in substantial variation in the percentages of household members reported as having long-term health conditions, disabilities and restrictive health conditions that limit their everyday activities as shown in Tables 41a and 41b (on page 78).

Table 41a: Any household members (other than the study child) with a disability, K cohort, Waves 1 to 6 (%)

	Age of study child					
	Age 4–5 (2004)	Age 6–7 (2006)	Age 8–9 (2008)	Age 10–11 (2010)	Age 12–13 (2012)	Age 14–15 (2014)
Household member has a disability	24.7	22.1	18.3	14.4	27.8	27.6
Household member is restricted in everyday activities	**37.9**	21.2	16.0	12.6	20.9	23.7
Household member has a disability and is restricted in everyday activities	7.9	8.2	**6.7**	**5.4**	11.0	12.1
No family member has a disability or restricted in everyday activities	55.7	69.8	75.6	80.4	67.3	65.4
n	**4,983**	**4,464**	**4,331**	**4,164**	**3,956**	**3,537**

Notes: Population weighted results. Columns do not total 100.0 as study child may have more than one household member with a health condition or disability. For Wave 1, measures of disability and restrictive health conditions are calculated by separating the conditions listed in the show cards into two groups, corresponding to those from Wave 2 onwards. (Note: the other option is to remove W1 from the analysis of restrictive health conditions throughout the report.)

Source: LSAC K cohort, Waves 1–6.

Table 41b: Household member (other than the study child) has a disability, K cohort, Waves 1–6 (%)

	Age of study child					
	Age 4–5 (2004)	Age 6–7 (2006)	Age 8–9 (2008)	Age 10–11 (2010)	Age 12–13 (2012)	Age 14–15 (2014)
Household member has a disability and is restricted in everyday activities	7.9	8.2	6.7	5.4	11.0	12.1
Mother	4.1	2.8	2.4	0.8	5.1	6.0
Father	2.1	2.3	1.9	2.6	3.0	3.1
Sibling	1.8	2.9	2.0	1.9	2.9	3.2
Grandparent	0.4	0.8	0.8	0.3	1.0	1.1
Other household member	-	0.2	0.2	-	0.3	0.4
n	**4,983**	**4,464**	**4,331**	**4,164**	**3,956**	**3,537**

Notes: Population weighted results. Columns do not total 100.0 as study child may have more than one household member with a health condition or disability. For Wave 1, measures of disability and restrictive health conditions are calculated by separating the conditions listed in the show cards into two groups, corresponding to those from Wave 2 onwards. (Note: the other option is to remove W1 from the analysis of restrictive health conditions throughout the report.)

Source: LSAC K cohort, Waves 1–6.

30 Academic Rating Scale score in Wave 7

The Longitudinal Study of Australian Children (LSAC) uses the Academic Rating Scale (ARS) as one measure of children's academic development. The ARS is also used in the Early Childhood Longitudinal Study (ECLS-K) in the United States (see National Center for Education Statistics [NCES], 2002, 2004). The ARS in the earlier years of ECLS-K is divided into three domains: Language and Literacy, Mathematical Thinking and General Knowledge.

The original ARS was adapted for use with Australian children for LSAC. Only the Language and Literacy and Mathematical Thinking domains were used with the K cohort in Waves 2 through 6, and with the B cohort in Waves 4 through 6. In Wave 7, only the Language and Literacy domain was used, and only with the B cohort.

This section describes the procedures followed to obtain scores for the Academic Rating Scale in Wave 7.

In LSAC, the ARS is administered as part of the Teacher Questionnaire. The nine Language and Literacy items in the questionnaire ask the study child's English teacher about the child's skills, knowledge and behaviours as evidenced in the child's current achievement and motivation, compared to other children in the same year level.[6] There are five levels of rating: Not yet, Beginning, In progress, Intermediate and Proficient. Teachers can also indicate if the skill has not yet been introduced at the year level.

30.1 Method

LSAC ARS scores were calculated in the same manner as the ARS scores in ECLS-K, using the Rasch rating score model. This is the procedure followed in previous waves of LSAC for both the K and B cohorts.

Only children who were rated on more than 60% of items were assigned rating scale scores. In Wave 7, the Language and Literacy domain comprised nine items; ratings were therefore required on six or more items. Children with scores on fewer items were not included in the analyses and were not assigned scores. The numbers of children who were and were not assigned scores in each Wave are contained in Table 42.

Table 42: Number of children assigned scores on the Academic Rating Scale, Language and Literacy, Wave 7

	Language and Literacy
ARS score assigned	2,524
Some items rated but no score assigned	29
No score on any item	828
Total	**3,381**

Scores on each of the nine items were rated from 1 to 5, according to the skill level assigned by the teacher. The initial analyses indicated that there was no overlap of the steps within each item (see Appendix A).

Principal component analysis indicated that a single component could be extracted from the nine items, accounting for 71.5 % of the variance (see Appendix B). Subsequent analysis used the rating scale option of the Rasch model, based on Wright and Masters (1982) and implemented in Quest (Adams & Khoo, 1996). The Rasch analysis showed that the reliability of the estimates of children's ability in Language and Literacy was very high (see Table 43). These estimates have remained above 0.90 for both cohorts, with decreases showing once the children have entered secondary school. The analysis assigned case estimates to each child.

6 It is assumed that all children in the cohort have commenced secondary education, with different teachers for each learning area.

Table 43: Internal consistency statistics for the Academic Rating Scale, Language and Literacy, by cohort and wave

Cohort	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7
K cohort	0.95	0.96	0.95	0.93	0.92	- -
B cohort	- -	- -	0.96	0.95	0.94	0.92

Perfect scores were estimated by adding 1.1 logits to the Rasch estimate of the second highest score. 'Zero' scores were estimated by subtracting 1.1 logits from the Rasch estimate of the second lowest score. The Rasch model does not calculate estimates for perfect or zero scores – 'extreme scores' – so some estimation is required. Wright (1998) has suggested that these extreme scores should be at least 1.0 logit and no more than 1.2 logits away from the next scores, unless some justification can be made for using a greater distance. Examination of the distances between the near-perfect and near-zero scores showed that the addition/subtraction of between 1.0 and 1.2 logits for an extreme score was appropriate, and that 1.1 logits would provide a reasonable result.

Once case estimates were obtained for the pattern of ratings on the ARS items, the estimates were transformed to ARS scores that reflect the range of scores available to the children's teachers; that is, the lowest possible score on the ARS scale is 1 and the highest is 5. Rasch case estimates were then transformed to ARS scores using a linear transformation. Again, this is consistent with the procedures used in ECLS-K. The equation used to convert the Rasch estimates to ARS scores is:

ARS = 2.9513 + (0.2784 x estimate)

Table 44 presents the conversion data for the ARS in each domain for children who obtained scores on all items in the scale. The table shows the raw score, the ARS score and the standard error associated with each score. As noted above, ARS scores were assigned to children with ratings on at least 60% of the items in a scale.

Table 44: Raw score to Academic Rating Scale score conversion tables, Language and Literacy, B cohort, Wave 7

Language and Literacy		
Raw score	ARSLIT score	Standard error (s.e.)
9E	1.00	--
10	1.31	0.30
11	1.54	0.23
12	1.69	0.20
13	1.81	0.18
14	1.91	0.17
15	2.00	0.17
16	2.09	0.16
17	2.17	0.16
18	2.24	0.16
19	2.32	0.15
20	2.39	0.15
21	2.47	0.15
22	2.54	0.15
23	2.61	0.16
24	2.69	0.16
25	2.76	0.16
26	2.83	0.16
27	2.91	0.16
28	2.98	0.16
29	3.06	0.16

Table continued on next page →

Language and Literacy		
Raw score	ARSLIT score	Standard error (s.e.)
30	3.14	0.16
31	3.22	0.16
32	3.30	0.17
33	3.39	0.17
34	3.47	0.17
35	3.56	0.17
36	3.65	0.18
37	3.75	0.18
38	3.84	0.18
39	3.95	0.18
40	4.05	0.18
41	4.16	0.19
42	4.30	0.21
43	4.46	0.24
44	4.69	0.31
45E	5.00	--

Note: Extreme scores are indicated with E; standard errors not available for extreme scores.

30.2 Results and using ARS scores

Summary statistics for the final ARS Language and Literacy scores are shown in Table 45, for all waves in which the ARS was administered with the B cohort. It should be noted that scores are determined independently for each wave, as each score is based on the teacher's evaluation of how well the student is achieving against his or her peers at the same year level.

Table 45: Summary statistics, Academic Rating Scale scores, Language and Literacy, B cohort, Waves 4–7

Wave	*n*	Mean	Standard deviation
4	3,408	3.40	0.772
5	3,455	3.59	0.837
6	3,087	3.83	0.815
7	2,524	3.85	0.786

The mean score in the Language and Literacy domain for each Wave in Table 45 indicates that, on average, members of the B cohort have been rated between 'In progress' and 'Intermediate'.

The ARS scores are measures that have been placed on an interval scale enabling comparisons between groups. Items in the ARS from earlier waves are not comparable to items in later waves. The following advice has been provided to users of ECLS-K data, and it applies to LSAC data:

> *The ARS scale was designed to provide information on children's abilities* ***at a given point in time, not necessarily over time****. In addition, although some item stems are similar to those used in the kindergarten and first grade teacher questionnaires, the actual items include performance criteria that increase in difficulty from one time to the next. Moreover, the ARS scores are placed on different metrics relative to the item difficulty in a given grade. Therefore,* ***change scores should not be calculated between time points****. However, covariance models may be used to compare teacher's ratings of performance in different grades. Before using these variables in such analyses, the distribution of the samples should be assessed to determine if the assumption of normal distribution is met. (NCES, 2004, p. 3–33; emphasis added)*

In the case of LSAC, the items used in Wave 7 were not changed from Wave 6; however, as noted above, the reference point in each Wave is other children in the same year level.

31 Gambling data inconsistencies

In Wave 7 the K cohort Parents and Study Children were asked about their participation in gambling activities over the past 12 months.

ise26a1a – ise26a1j (Parent 1)/ ise26c1a - ise26c1j (Study Child) – "During the last 12 months have you spent money on any of the following (Scratchies, Bingo, Lotto, Keno, Private betting, Poker, Casino games, Poker machines, Horse or dog races, Sports betting)?" **Yes or No**

For **"Yes"** responses to above question, participants were then asked about the not online and online frequency of the gambling activity they had spent money on.

ise26a2a – ise26a2j (Parent 1)
ise26c2a – ise26c2j (Study Child)
Thinking about the past 12 months how often have you participated in "*Word Sub gambling activities specified in previous question*" **not online**?

ise26a3a – ise26a3j (Parent 1)
ise26c3a – ise26c3j (Study Child)
Thinking about the past 12 months how often have you participated in "*Word Sub gambling activities specified in previous question*" **online**?

There is some inconsistency in the data when comparing the **"Yes"** responses to the top question (ise26a1(P1) / ise26c1(SC)) with the subsequent questions regarding frequency of not online and online gambling. 137 parent records and 112 Study Child records have have indicated at the top level question that they had spent money on particular gambling activities in the past 12 months, but then in the subsequent questions about the frequency of this activity in the past 12 months they have answered "never" or "not in the past twelve months" to both not online and online, which is inconsistent with their response to the first question. As a result the variables for "method gambled in last 12 months" ise26a5 (P1) / ise26c5 (SC) have been removed from release 7.1 as they were derived from the not online and online frequencies. There is also a difference in the frequency values between the parent and study child items. Parents had a value range of 0 to 6 whereas the Study Child had a value range of 0 to 8.

Users should be cautious when working with the not online and online frequency data and be aware that some inconsistencies may be highlighted if data is compared to the other gambling questions.

32 Income imputation and household income derivation

Not all input variables to impute Parent 1 and Parent 2 income are available in Wave 8. Some measures have either not been asked or cannot be rolled forward from what is available. It is therefore not possible to output Parent 1, Parent 2, Mother, Father and Household imputed income variables. Imputation of income measures has been done in previous waves using the Nearest Neighbour and Little and Su methods (Mullan, Daraganova, & Baker, 2015)).

The following variables that were previously used in the imputation process, but are no longer being collected or rolled forward from Wave 8 K cohort are:

- What is the highest year of primary or secondary school you completed [Parent 1 and Parent 2]?
- Housing tenure
- Parent 1 and Parent 2 work status
- Income from Other adults.

Until Wave 7, household income was computed by adding together Parent 1, Parent 2 (if with P1) and any other adults in the house 15 years or older for B and K cohorts. The income of other adults in the house was still asked of the Wave 8 B cohort and discontinued for Wave 8 K cohort. Therefore, household income will not be provided from Wave 8 onwards for the K cohort, and this measure will still be provided for Wave 8 B cohort. Personal income variables for Parent 1, Parent 2, Mother, Father, Young people and their Partners are available in the data files.

It is worth noting that Young people's and their Partners' incomes are not imputed using prior work-related variables because a majority of them had not entered into the workforce.

33 Household Socio-economic positioning (SEP)

LSAC has derived a measure of socio-economic position (SEP) based on the education, income and employment characteristics of Parent 1 and Parent 2. It was designed as a measure of household/family socio-economic position that could be used in research looking at child outcomes. However, the ongoing calculation of parental household socio-economic status is less relevant as children enter adulthood from Wave 8 onwards. At Wave 8, 15–20% of young people no longer lived with P1. Many are financially independent of their parents, and some of them have their own jobs and/or partners with jobs. While the characteristics of the household(s) they grew up in continue to be relevant to ongoing outcomes, the current parental household characteristics are less relevant to current and future outcomes. It has also become progressively more problematic to produce parental SEP measures from Wave 8 onwards because some of the key characteristics, such as parental education, are no longer being collected and updated. Therefore, parental SEP measures will not output for Wave 8 onwards for the K cohort. This measure will still be provided for the Wave 8 B cohort.

34 Parent's childhood experiences – differences between cohorts for breadwinner questions

Since Wave 4 parents have been asked questions about their childhood experiences. They were asked to think back to when they were 14 years old. Included in this construct were questions about their parent's occupation when the Parent 1 or Parent 2 was 14 years old.

The first approach taken in Wave 4 and 5 was to ask occupation questions about the main breadwinner for both cohorts. In Wave 7 the approach was to then ask about the other parent (i.e. the non-breadwinner). However, previously it has been noted that non-breadwinner data were dropped from Wave 7 due to data quality issues for both cohorts.

In Wave 8 these non-breadwinner questions were asked in the B cohort parent interview. However, for the K cohort parents, with the change in methodology to a shorter telephone interview only, these questions were not included in the final content.

Therefore, within the pa24 topic, data users should expect to find longitudinal data that are more comprehensive for the B cohort parents as compared to the K cohort parents.

35 Event History Calendar (EHC) issues

Due to an instrument error in the Wave 8 EHC module, the data for the item 'Reasons took time out from study' (variables: jehcs15a - jehcs15r, jtakoth) are not available for Wave 8.

For a small number of records (5) there will be an inconsistency between EHC information and data collected during the main interview. More specifically, there will be no EHC work or study episodic data available; however, there will be some work and study information available within the main dataset.

36 Missing data from online component of Wave 8

In the early stages of Wave 8 enumeration there were 55 Main Wave K cohort participants who completed an incorrect version of the online instrument, due to two versions being available. As a result, there will be some missing data for these records, given some content was missing from this early version.

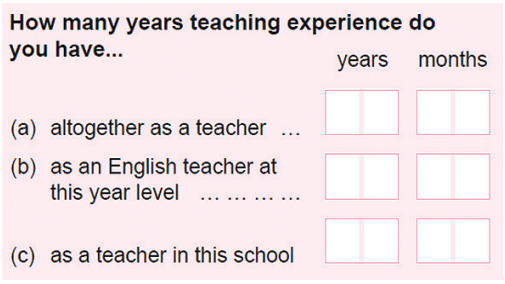
37 Job Security

In Wave 7 the raw data item for the Parent 2 variable gpw21b (How secure do you feel in your present job?) was not reverse coded. The table below shows the data before and after this correction.

gpw21b	Before correction	After correction
-9	1,467	1,467
1 – Very insecure	688	42
2 – Not very secure	861	243
3 – Secure	243	861
4 – Very secure	42	688

38 Teacher Experience

In Wave 7 B cohort, teachers were asked about teaching experience and data were collected in years and months. In addition, the data are then also output as total year and total months, combining the two input items.



For the Wave 7 release there were two issues in the data for teaching experience for (b) as an English teacher at this year level and (c) as a teacher in this school.

1. There was an error in the derive for total years and months due to incorrect use of rounding code.
2. There was also a mapping issue where the collected years and months for (b) and (c) were shown on the released final file as the derived total years and total months and vice versa.

The variables affected were:

gpc32c1 – years teaching experience? as an English teacher at this grade level? Total years derived

gpc32c2 – years teaching experience? as an English teacher at this grade level? Total months derived

gpc32c2a – years teaching experience? as an English teacher at this grade level? Years part

gpc32c2b – years teaching experience? as an English teacher at this grade level? Months part

gpc32b2a – years teaching experience? as a teacher in this school? Total years derived

gpc32b2b – years teaching experience? as a teacher in this school? Total months derived

gpc32b2c – years teaching experience? as a teacher in this school? Years part

gpc32b2d – years teaching experience? as a teacher in this school? Months part

39 Location of most serious injury

In Wave 7, the Parent 1 was asked about where the study child's most serious injury in the last 12 months had occurred (K Cohort ihs18a15a2 and B Cohort ghs18e2). The response options for the Parent 1 were slightly different between the cohorts. The raw items were not renumbered to match the output categories. The table below shows the data before and after this correction.

ghs18e2	Before correction	After correction
-9	2,519	2,519
10	123	
11	247	
12	48	
13	307	
14	95	
15	14	
16	28	
1 – At home		123
2 – School or child care		247
3 – Someone else's place e.g. family member/friends/neighbours		48
4 – Outside public place other than a road e.g. beach, playground, sports ground		307
5 – Inside public place e.g. shopping centre, gym, indoor sports centre		95
6 – Public road		14
7 – Other		28

ihs18a15a2	Before correction	After correction
-9	2,273	2,273
10	87	
11	163	
12	28	
13	28	
14	385	
15	83	
16	18	
17	24	
1 – At home		87
2 – School		163
3 – Work		28
4 – Someone else's place e.g. family member/friends/neighbours		28
5 – Outside public place other than a road e.g. beach, playground, sports ground		385
6 – Inside public place e.g. shopping centre, gym, indoor sports centre		83
7 – Public road		18
8 – Other		24

40 Personality data missing

In Wave 7, Parent 2 K cohort, for the personality item 'tends to be lazy' there was missing data for 1,125 records. The table below shows the data before and after this correction.

ise30b3	Before correction	After correction
-9	1,309	1,309
1 – Disagree strongly	308	802
2 – Disagree a little	132	370
3 – Neither agree nor disagree	89	267
4 – Agree a little	102	294
5 – Agree strongly	13	36
Missing	1,136	11

41 Changes to 'Consent to contact Parent Living Elsewhere (PLE)' variables

From Wave 2 to Wave 3, variables *id29 related to a direct question asked of the P1 to gain consent to interview the Parent Living Elsewhere (PLE).

For Wave 4, no PLE consent to contact variables were added to the dataset.

From Wave 5 onwards, there was a change in the approach to collecting permission to contact the PLE. The permission to contact the PLE was collected (still from P1) in a more passive way through obtaining phone numbers for the PLE. The variable names and question details in the Data Dictionary released in Wave 5 through to Wave 7 did not accurately reflect this change. *id29 was used as an indicator to show if the P1 had entered the PLE module (or not) and a new variable *id29a was introduced to reflect the collection of passive permission to contact the PLE. To assist users in understanding these variables better the following changes have been made:

Summary of changes made:

- hid29 (PLE entered the PLE module) dropped for Wave 8 B cohort (not considered to be a useful item to data users anymore).
- Variable label and question text updated for Waves 5, 6, 7 for *id29 to indicate that this variable was showing where P1 has entered the PLE module.
- Variable label and question text updated for Waves 5, 6, 7 for *id29a to highlight that this was derived as passive consent through provision of contact details for PLE.

42 Academic Rating Scale score in Wave 8

The Longitudinal Study of Australian Children (LSAC) uses the Academic Rating Scale (ARS) as one measure of children's academic development. The ARS has been used in the Early Childhood Longitudinal Study (ECLS-K) in the United States (see NCES, 2002, 2004) in relation to study children in Years 1, 3 and 5. The ARS in the earlier years of ECLS-K is divided into three domains: Language and Literacy, Mathematical Thinking and General Knowledge. For children in secondary school (Year 8), the ECLS-K uses separate questionnaires with teachers of English, Mathematics and Science.

The original ARS was adapted for use with Australian children for LSAC. Only the Language and Literacy and Mathematical Thinking domains were used with the K cohort in Waves 2 through 6, and with the B cohort in Waves 4 through 6. In Waves 7 and 8, only the Language and Literacy domain has been used, and only with the B cohort, as many of the children in the K cohort had left school.

This paper describes the procedures followed to obtain scores for the Academic Rating Scale in Wave 8. The procedures remain unchanged since the ARS was first used in LSAC.

In LSAC, the ARS is administered as part of the Teacher Questionnaire. The nine Language and Literacy items in the questionnaire ask the study child's English teacher, or another teacher who has regular classroom contact with the child, about the child's skills, knowledge and behaviours as shown in the child's current achievement and motivation, compared to other children in the same year level. There are five levels of rating: Not yet, Beginning, In progress, Intermediate and Proficient. Teachers can also indicate if the skill has not yet been introduced at the year level[7].

42.1 Method

LSAC ARS scores were calculated in the same manner as the ARS scores in ECLS-K, using the Rasch rating score model. This is the procedure followed in previous waves of LSAC for both the K and B cohorts.

Only children who were rated on more than 60% of items were assigned rating scale scores. In Wave 8, the Language and Literacy domain comprised nine items; ratings were therefore required on six or more items. Children with scores on fewer items were not included in the analyses and were not assigned scores. The numbers of children who were and were not assigned scores in each Wave are contained in Table 48.

Possible scores on each of the nine items are 1, 2, 3, 4 or 5, according to the skill level assigned by the teacher. The initial analyses indicated that there was no overlap of the steps within each item (see Appendix C).

7 It is assumed that all children in the cohort have commenced secondary education, with different teachers for each learning area.

Table 48: Number of children assigned scores on the Academic Rating Scale, Language and Literacy, Wave 8

	Language and Literacy
ARS score assigned	2,290
Some items rated but no score assigned	12
No score on any item	825
Total	**3,127**

Principal component analysis indicated that a single component could be extracted from the nine items, accounting for 76.9% of the variance (see Appendix D). Subsequent analysis used the rating scale option of the Rasch model, based on Wright and Masters (1982) and implemented in Quest (Adams & Khoo, 1996). The Rasch analysis showed that the reliability of the estimates of children's ability in Language and Literacy was very high (see Table 49). These estimates have remained above 0.90 for both cohorts, with decreases showing once the children have entered secondary school. The analysis assigned case estimates to each child.

Table 49: Internal consistency statistics for the Academic Rating Scale, Language and Literacy, by cohort and wave

Cohort	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8
K cohort	0.95	0.96	0.95	0.93	0.92	- -	- -
B cohort	- -	- -	0.96	0.95	0.94	0.92	0.92

Perfect scores were estimated by adding 1.1 logits to the Rasch estimate of the second highest score. 'Zero' scores were estimated by subtracting 1.1 logits from the Rasch estimate of the second lowest score. The Rasch model does not calculate estimates for perfect or zero scores – 'extreme scores' – so some estimation is required. Wright (1998) has suggested that these extreme scores should be at least 1.0 logit and no more than 1.2 logits away from the next scores, unless some justification can be made for using a greater distance. Examination of the distances between the near-perfect and near-zero scores showed that the addition/subtraction of between 1.0 and 1.2 logits for an extreme score was appropriate, and that 1.1 logits would provide a reasonable result.

Once case estimates were obtained for the pattern of ratings on the ARS items, the estimates were transformed to ARS scores that reflect the range of scores available to the children's teachers; that is, the lowest possible score on the ARS scale is 1 and the highest is 5. Rasch case estimates were then transformed to ARS scores using a linear transformation. Again, this is consistent with the procedures used in ECLS-K. The equation used to convert the Rasch estimates to ARS scores is:

$$ARS = 2.9065 + (0.2876 \times estimate)$$

Table 50 presents the conversion data for the ARS in each domain for children who obtained scores on all items in the scale. The table shows the raw score, the ARS score and the standard error associated with each score. As noted above, ARS scores were assigned to children with ratings on at least 60% of the items in a scale.

Table 50: Raw score to Academic Rating Scale score conversion tables, Language and Literacy, B cohort, Wave 8

	Language and Literacy	
Raw score	ARSLIT score	Standard error (s.e.)
9E	1.00	--
10	1.32	0.30
11	1.55	0.23
12	1.69	0.19
13	1.81	0.17
14	1.91	0.16
15	1.99	0.15
16	2.07	0.15
17	2.14	0.14

Table continued on next page →

Language and Literacy		
Raw score	ARSLIT score	Standard error (s.e.)
18	2.21	0.14
19	2.28	0.14
20	2.34	0.14
21	2.41	0.14
22	2.47	0.14
23	2.54	0.14
24	2.61	0.14
25	2.68	0.14
26	2.75	0.14
27	2.82	0.15
28	2.90	0.15
29	2.97	0.15
30	3.05	0.15
31	3.14	0.16
32	3.22	0.16
33	3.31	0.16
34	3.40	0.16
35	3.49	0.17
36	3.59	0.17
37	3.69	0.17
38	3.79	0.17
39	3.90	0.18
40	4.01	0.18
41	4.13	0.19
42	4.27	0.21
43	4.44	0.24
44	4.68	0.31
45E	5.00	--

Notes: Extreme scores are indicated with E; standard errors not available for extreme scores.

42.2 Results and using ARS scores

Summary statistics for the final ARS Language and Literacy scores are shown in Table 51, for all waves in which the ARS was administered with the B cohort. Scores are determined independently for each wave, as each score is based on the teacher's evaluation of how well the student is achieving against his or her peers at the same year level.

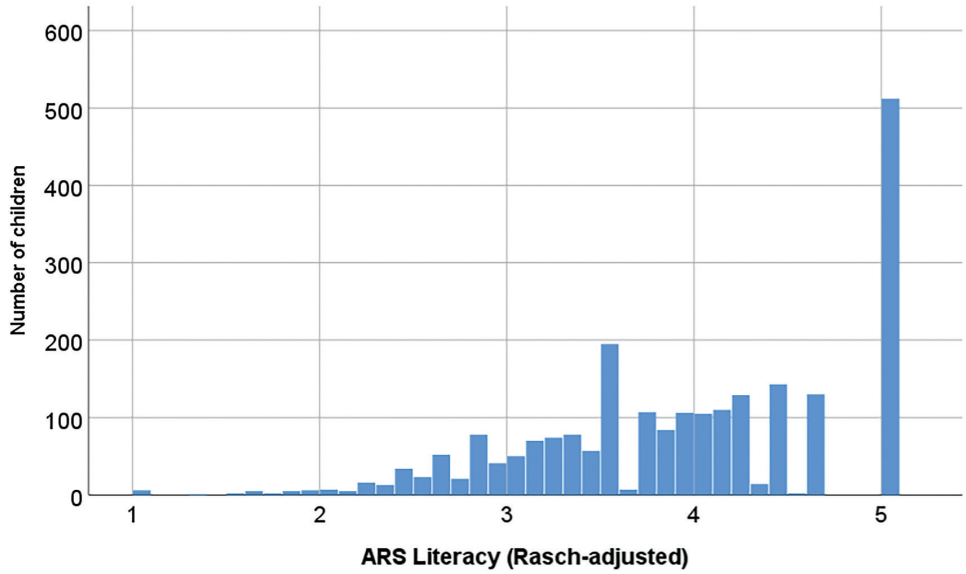
Table 51: Summary statistics, Academic Rating Scale scores, Language and Literacy, B cohort, Waves 4–8

Wave	*N*	Mean	Standard deviation
4	3,408	3.40	0.772
5	3,455	3.59	0.837
6	3,087	3.83	0.815
7	2,524	3.85	0.786
8	2,290	3.95	0.823

The mean score in the Language and Literacy domain for each Wave in Table 51 indicates that, on average, members of the B cohort have been rated between 'In progress' and 'Intermediate'.

Distribution of the ARS scores in Language and Literacy are highly skewed, with 512 children (22% of children receiving scores) receiving rating scale scores of 5 across all items on the scale (Figure 8). One-half of all children who received scores were rated at 4 or above.

Figure 8: Distribution of Academic Rating Scale scores, Language and Literacy, B cohort, Wave 8



The ARS is a measure on an interval scale enabling comparisons between groups. Items in the ARS from earlier waves are not comparable to items in later waves. The following advice has been provided to users of ECLS-K data, and it applies to LSAC data:

> *The ARS scale was designed to provide information on children's abilities* ***at a given point in time, not necessarily over time****. In addition, although some item stems are similar to those used in the kindergarten and first grade teacher questionnaires, the actual items include performance criteria that increase in difficulty from one time to the next. Moreover, the ARS scores are placed on different metrics relative to the item difficulty in a given grade. Therefore,* ***change scores should not be calculated between time points****. However, covariance models may be used to compare teacher's ratings of performance in different grades. Before using these variables in such analyses, the distribution of the samples should be assessed to determine if the assumption of normal distribution is met. (NCES, 2004, p. 3-33; emphasis added)*

In the case of LSAC, the items used in Wave 8 were not changed from previous waves; however, as noted above, the reference point in each Wave is *other children in the same year level*.

We also note that scores on the ARS are highly skewed, with one-half of children receiving a score of 4 or above, and that the ARS is no longer used as part of the ECLS-K in the United States for children who have reached secondary school. Lower scores on the ARS may provide a good indication of children who is experiencing difficulties in school, but the skewed distribution makes it difficult to discriminate among children with higher scores.

43 Relationship of all members to Young Person – comparison of Wave 7 and Wave 8

In Wave 8, the Young Person (now 18–19 years of age) was approached directly for an interview. They also completed the HHF module to confirm who they were living with, when previously the Parent 1 completed this module, as the best person to report on behalf of the Young Person.

During this module, the Young Person responded to questions about household changes based on pre-filled information about the family they lived with in the previous wave. However, they were required to report from scratch what their relationship was to each family member they lived with in Wave 8.

The outcome of the methodological change was that, for some records, the Young Person reported a different relationship to what Parent 1 had previously reported in Wave 7 or earlier waves. In some cases, the relationship change was impossible or highly unlikely and, given it was assumed the response was an error, the data were amended to the Wave 7 (or earlier) response. However, for the majority of cases, the information was left in Wave 8 as reported by the Young Person. Data users of the household file may, therefore, find extra differences between the items in Wave 7 (if08m2—if08m21) and Wave 8 (jf08m2—jf08m21).

44 Imputations to solve missing data problems in items for Number of People in the Household in each age bracket

A number of variables in Wave 9C1 have missing data more than what is usual for an LSAC dataset. This is due to respondents skipping or not answering the questions. This section outlines the imputations made in order to limit the amount of missing data in the 'number of people in the household in each age bracket' items that were asked of the P1, P2 and PLE. Note: The Study Child was also asked about the number of people they lived with, but they were not asked for a breakdown of the household members by age bracket so this missing data issue did not occur for their data.

As a reference, Section 5.8 from Data User Guide contains the list of missing convention used for LSAC. A new code frame value of -5 has been introduced in the list to refer to web form questions that has been missed and/or skipped by respondents and this has been used to treat the missing data accordingly.

In the introduction module of the CAWI, P1, P2 and PLE respondents were asked about who lives with them and then were asked about the number of people in each age bracket:

INTRO_Q07 (i1fd32a, i1fd32b, i1fd32p, k1fd32a, k1fd32b, k1fd32p)

The next questions are about who lives with you.

Apart from Study Child if they lived with them, how many people currently live with you?

Include household members who live with you some of the time.

INTRO_Q09 (i1fd33a1a - i1fd33a5a, i1fd33b1a - i1fd33b5a, i1fd33p1a - i1fd33p5a and k1fd33a1a - k1fd33a5a, k1fd33b1a - k1fd33b5a, k1fd33p1a - k1fd33p5a)

How many people in your household (excluding you and Study Child if respondent lives with Study Child, otherwise, excluding the only respondent) are in each of the following age groups?

If you do not know the age of someone you live with, your best guess is fine.

1. Under 5
2. 5–12
3. 13–17
4. 18–64
5. 65 or over

The variables associated with the number of people in the household in each of the age groups have a higher proportion of missing data than other items. The respondents were sequenced to the items but they didn't enter a response, or only partially responded. An example of one of the age bracket items with missing data, and the corrections made, is as follows:

I1fd33a1a - 16/17 - P1 - P CAWI A2.3.1 - How many HH people are in the age group Under 5

Table 52 shows just for P1 B cohort records the number of people in the HH in the age group under 5 prior to any imputation and prior to any treatment of missing data using -5 values.

Table 52: Before imputing 0 and before treatment of missing web form data

16/17 - P1 - P CAWI A2.3.1 - How many HH people are in the age group Under 5				
i1fd33a1a	Frequency	%	Cumulative frequency	Cumulative %
-9	770	91.13	770	91.13
0	43	5.09	813	96.22
1	26	3.08	839	99.29
2	5	0.60	844	99.89
3	1	0.12	845	100.00
Frequency missing = 1,172				

The assumption that has been made here is that the majority of the missing records (1,172 in the table above) should have selected 0 people for the age bracket but instead didn't provide a response because they didn't think they had to select 0 people.

As a result, the response value of '0 people' has been imputed for all of these age bracket variables if the following conditions were met:

- if one or more of the age group brackets was answered

if the total number of people (i.e. when adding up the five age brackets) is equal to the number of people in the house (excluding parent respondent and YP) as answered in the earlier question INTRO07 (i1fd32a, i1fd32b, i1fd32p, k1fd32a, k1fd32b, k1fd32p) After implementing this amendment the number of records showing 0 people in the under 5 age bracket has increased to 926. The number of missing records for this item has reduced to 289 (which have now been changed to -5 as per new missing web value conventions). These 289 records were not able to be imputed because if we impute missing values to 0 then the total of the age groups does not match the value provided in INTRO07.

Table 53: After imputing 0 and after treatment of missing web form data

16/17 - P1 - P CAWI A2.3.1 - How many HH people are in the age group Under 5				
i1fd33a1a	Frequency	%	Cumulative frequency	Cumulative %
-9	770	38.18	770	38.18
-5	289	14.33	1,059	52.50
0	926	45.91	1,985	98.41
1	26	1.29	2,011	99.70
2	5	0.25	2,016	99.95
3	1	0.05	2,017	100.00

45 Explanation of some not applicable (-9) data in 'Have you ever had even part of an alcoholic drink' (Wave 7 Compound item) ihb16c11a

There are 1,059 K SC records that provided a valid response of 'yes, a few sips' to the 'ever had alcohol question' (hhb16c11) in Wave 6. It's important to note that those respondents just having a few sips were not to be counted as 'ever drunk alcohol' for the purposes of the compound item (hb16c11a). These records that answered 'yes, a few sips' were inadvertently given an incorrect Wave 7 pre-fill value that meant that they weren't asked the 'ever had alcohol' question again in Wave 7 when they should have been. As a result, the Wave 7 compound item (ihb16c11a) is showing -9 for these records. All other records were correctly pre-filled and the Wave 7 compound item accurately reflects if they have 'ever drunk alcohol'.

The affected population did have the opportunity to answer the following questions in Wave 7:

- Have you had an alcoholic drink in the last 12 months? (ihb16c13)
- Have you had an alcoholic drink in the last four weeks? (ihb16c9)
- The number of alcoholic drinks you had during the last seven days, including yesterday.

These data could possibly be used to populate the Yes category in the 'ever drunk alcohol' Wave 7 compound item. However, the risk here is that unless the respondent answered the last question (number of drinks in the last seven days), we won't know if it was just a few sips or a full drink. So the decision was made to not do this amendment.

These affected records will have been asked the 'ever drunk alcohol' question again in Wave 8 (jhb16c11) and the Wave 8 compound item (jhb16c11a) will reflect their Wave 8 response.

46 ANZSCO coding for Study Child desired future occupation data items

The desired future occupation question has been asked of the Study Child or Young Person from Wave 6 through to Wave 8. The variables from Waves 6–8 are summarised in the table below.

Wave	Cohort	Variable name	Question ID	Variable label	Question
6	K	hpw39ca1	pw39_a1	14/15 - CSRK 13.2 - Child's desired occupation	What is your desired occupation?
7	K	ipw39ca1	pw39_a1	16/17 - SC/RAP - SC CAI B6.8 - Desired occupation	What is your desired occupation?
8	B	hpw39ca1	pw39_a1	14/15 - CSRB 10.2 - Child's desired occupation	What is your desired occupation?
8	K	jpw39ca1	pw39_a1	18/19 - SC CASI F1.2 - Desired occupation	What is your desired occupation?
8	K	jpw44c2	pw44_2	18/19 - SC CASI F3.2 - 30 years old - Work type expectation	What kind of work do you expect to be doing when you are 30 years old?

These items were asked using a free text field response option. In Wave 8, the K cohort had the additional question: 'What kind of work do you expect to be doing when 30 years old?' The lead-in question for this item was:

'When you are 30 years old, do you think you will be ...' 1. Working full-time, 2. Working part-time, 3. Not working, or this item could be refused to be answered (Ctrl R). Without the specific question 'Do you know what career or occupation you would like to have in the future?' the number of 'Don't know' or nonsense responses was much higher for this item. (There were 222 responses of 'Don't know' for jpw44c2 compared to only 1–2 for the other future occupation variables.)

Up until the release of Wave 9C1 data the responses have only been made available on the LSAC restricted release files. To be able to add the data to the general release files and therefore make the data more accessible to data users, cleaning of the responses was undertaken, and responses were then coded to the ANZSCO classification. This is now in line with other similar occupation-style items in the LSAC datasets. In many cases more than one occupation was listed by the respondent and these additional occupations have also been coded. The responses have been coded to 4-digit occupation level ANZSCO code.

Some responses provided were unable to be coded and a very small number of responses are outside of the labour force. See the table below for a full description of these scenarios.

ANZSCO code	Description
0998	Unable to be coded or inadequately described – can't be coded to even the broadest, 1-digit level of ANZSCO (e.g. the likes of 'working with animals' – which could be a farmer, veterinarian, zoologist, veterinary nurse, farm hand)
0999	Outside of the labour force (e.g. a full-time parenting role)
-2	Don't know

These are the new items added to the data dictionary and final datasets as a result of this occupational coding work.

Variable Name	Variable Label
hpw39ca2a	14/15 - CSRK 13.2 - Child's desired occupation - ANZSCO - 1st code
hpw39ca2b	14/15 - CSRK 13.2 - Child's desired occupation - ANZSCO - 2nd code
hpw39ca2c	14/15 - CSRK 13.2 - Child's desired occupation - ANZSCO - 3rd code
hpw39ca2d	14/15 - CSRK 13.2 - Child's desired occupation - ANZSCO - 4th code
ipw39ca2a	16/17 - SC/RAP - SC CAI B6.8 - Desired occupation - ANZSCO - 1st code
ipw39ca2b	16/17 - SC/RAP - SC CAI B6.8 - Desired occupation - ANZSCO - 2nd code
ipw39ca2c	16/17 - SC/RAP - SC CAI B6.8 - Desired occupation - ANZSCO - 3rd code
hpw39ca2a	14/15 - CSRB 10.2 - Child's desired occupation - ANZSCO - 1st code
hpw39ca2b	14/15 - CSRB 10.2 - Child's desired occupation - ANZSCO - 2nd code
hpw39ca2c	14/15 - CSRB 10.2 - Child's desired occupation - ANZSCO - 3rd code
hpw39ca2d	14/15 - CSRB 10.2 - Child's desired occupation - ANZSCO - 4th code
jpw39ca2a	18/19 - SC CASI F1.2 - Desired occupation - ANZSCO - 1st code
jpw39ca2b	18/19 - SC CASI F1.2 - Desired occupation - ANZSCO - 2nd code
jpw39ca2c	18/19 - SC CASI F1.2 - Desired occupation - ANZSCO - 3rd code
jpw39ca2d	18/19 - SC CASI F1.2 - Desired occupation - ANZSCO - 4th code
jpw44c2aa	18/19 - SC CASI F3.2 - 30 yrs old - Work type expectation - ANZSCO - 1st code
jpw44c2ab	18/19 - SC CASI F3.2 - 30 yrs old - Work type expectation - ANZSCO - 2nd code
jpw44c2ac	18/19 - SC CASI F3.2 - 30 yrs old - Work type expectation - ANZSCO - 3rd code
jpw44c2ad	18/19 - SC CASI F3.2 - 30 yrs old - Work type expectation - ANZSCO - 4th code

47 Education items dropped in Wave 9C

In survey 9C1, no K cohort (age 20–21) young persons responded 'Secondary school' for question Educ_Q03 (What type of institute are you currently studying in?). Due to this, no K cohort young persons were sequenced through to Educ_Q15: 'Thinking about the year immediately after you leave school, what do you plan on doing?' (Population – Still in secondary school).

In Survey 9C2, 'What grade are you currently in now?' (W9C2 Educ_03c)

Grade/Year level currently in for K cohort (age 21–22) has no data due to K cohort no longer being in secondary school. This is the only variable with the population 'in secondary school'.

As a result, these K cohort items were dropped from the data dictionary and data files.

48 Number of people in the household in each age bracket (9C2)

Due to an instrument programming error, the responses to items (i/k2fd33a1a-i/k2fd33a5a and i/k2fd33c1-i/k2fd33c3) in Wave 9C, Survey 9C2 were all overridden to 0 after respondents had entered data. The initially entered data were able to be retrieved. For approximately 30 HICIDs, multiple entries were made by the respondent and their data were found to be unreliable and was set to missing '.'.

49 Sequencing error affecting Education items in 9C2

There were 370 B cohort SC records where they said they were currently enrolled in secondary school (i2pc82c2) but had also completed Year 12 (i2fd08c1a). At the first instrument sequencing point (see snippet below) these respondents should go down path 1 – those who have completed secondary school and path 3 – those who have completed Year 12, which is contradictory.

The respondents have been sequenced down path 1 as this is the first option. Using responses to other questions we have deduced that these respondents are enrolled in secondary school and have not yet completed Year 12. The response to i2fd08c1a has been set to missing '.'.

All	Q03a	**EDUC_SG1a** **1. Respondent is currently studying in secondary school (Q03=1)** **2. Respondent is not in secondary school (Q02=1 OR Q03 NE 1) AND did not complete Year 12 (Q03a NE 1)** **3. Otherwise (All who completed Year 12)**	 1 → Q03c 2 → Q03b 3 → Q03d

50 Comparability of Parent work items across 9C1 and 9C2

There were comparability issues for items measuring the effects on employment of someone studying at home during the coronavirus restriction period between March and May 2020 (CRP) across two surveys of Wave 9C (9C1 and 9C2). Therefore, a catchup item was not derived in 9C2 for the population who didn't submit their responses in 9C1.

1. There is the potential for the 9C2 respondents to report based on a different reference period compared to 9C1; that is, in 9C1, it would have been clear to respondents that it was talking about the first main lockdown period from March to May 2020.
2. There is a wider timespan of being an employee or self-employed in the 9C2 question compared to 9C1.

In survey 9C2, WORK_Q18 (P CAWI D8.4) and WORK_Q19 (P CAWI D 9.1) had a population of 'Did not submit 9C1', which is how catchup items are usually identified in the data processing. These two items had an additional population containing a reference period (Employee since March 2020). This was not consistent with the 9C1 population for this item and, therefore, a catchup item could not be created, and the items were just output as collected with the population as described in the Wave 9C2 survey.

References

ABS. (2006). *Australian and New Zealand Standard Classification of Occupations* (1st ed.). Cat. no. 1220.0. Canberra: ABS.

ABS. (2011). *Australian Statistical Geography Standard (ASGS): Volume 1. Main structure and greater capital city statistical areas*. Cat. no. 1270.0.55.001. Canberra: ABS.

de Lemos, M. & Doig, B. (2000). Who Am I? Supplementary Information. Melbourne: ACER.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test* (3rd ed.). Circle Pines, MN: American Guidance Service.

Egerton, M., & Gershuny, J. (2004). U*tility of time use data: Report to DfES*. Colchester: Institute for Social and Economic Research, University of Essex.

Fisher, K. (2002). *Chewing the fat: the story time diaries tell about physical activity in the United Kingdom*. Colchester: Institute for Social and Economic Research, University of Essex.

McMillan, J., Beavis, A., & Jones, F. L. (2009). The AUSEI06: A new socio-economic index for Australia. *Journal of Sociology, 45*(2), 123–149.

Mullan, K., Daraganova, G., & Baker, K. (2015). *Imputing income in the Longitudinal Study of Australian Children* (LSAC Technical Paper No. 14). Melbourne: Australian Institute of Family Studies. Retrieved from www.growingupinaustralia.gov.au/pubs/technical/tp14.pdf

National Center for Education Statistics (NCES). (2002). U*ser's manual for the ECLS-K first grade public-use data files and electronic code book* (NCES 2002-135). Washington, DC: US Department of Education, Office of Educational Research and Improvement.

National Center for Education Statistics (NCES). (2004). *User's manual for the ECLS-K third grade public-use data file and electronic code book* (NCES 2004-001). Washington, DC: US Department of Education, Institute of Education Sciences.

NCES. (2004). *User's manual for the ECLS-K third grade public-use data file and electronic code book* (NCES 2004-001). Washington, DC: US Department of Education, Institute of Education Sciences.

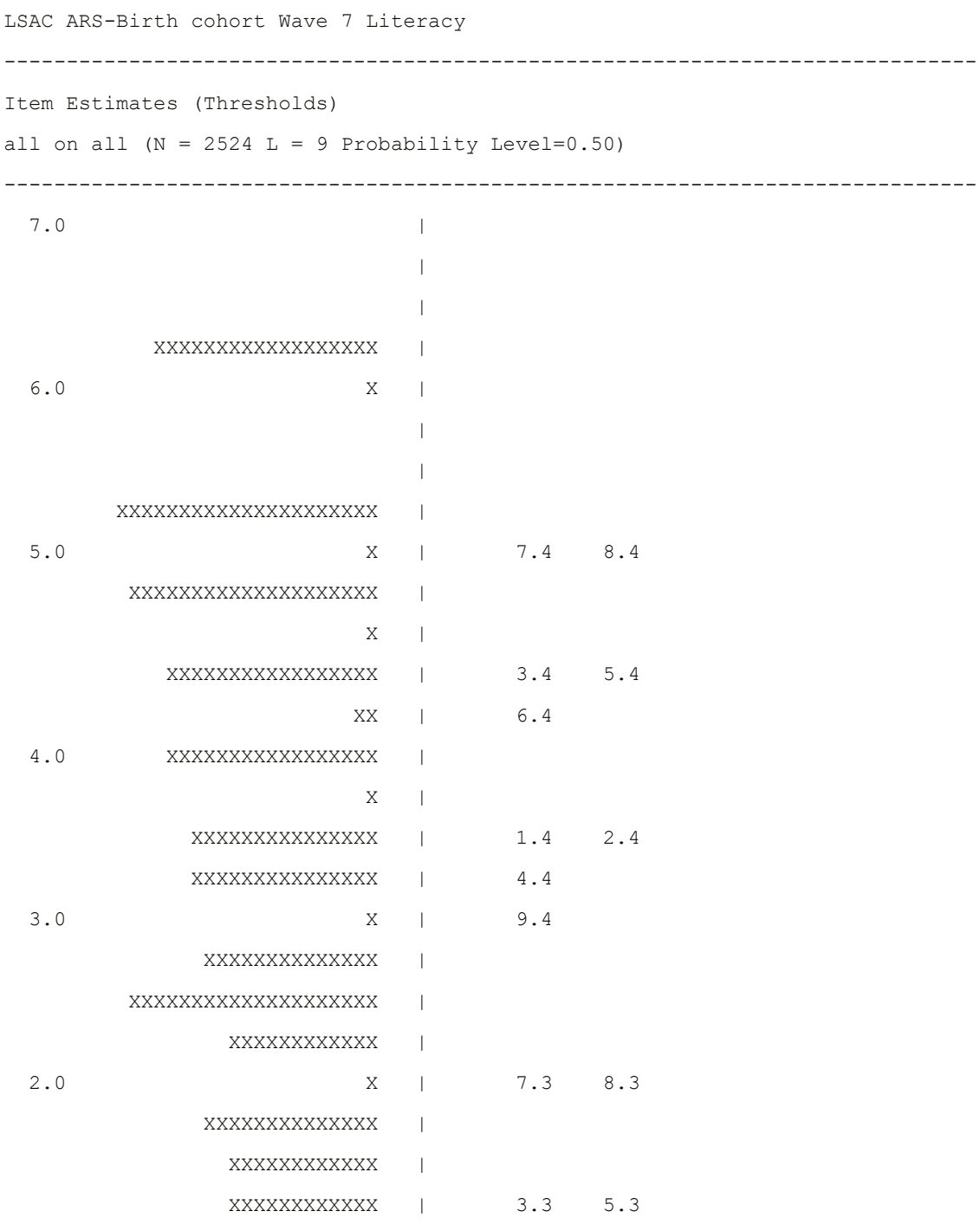
Rothman, S. (2013). *Using the Adapted PPVT-III in LSAC*. Unpublished paper. Melbourne: Australian Council for Educational Research.

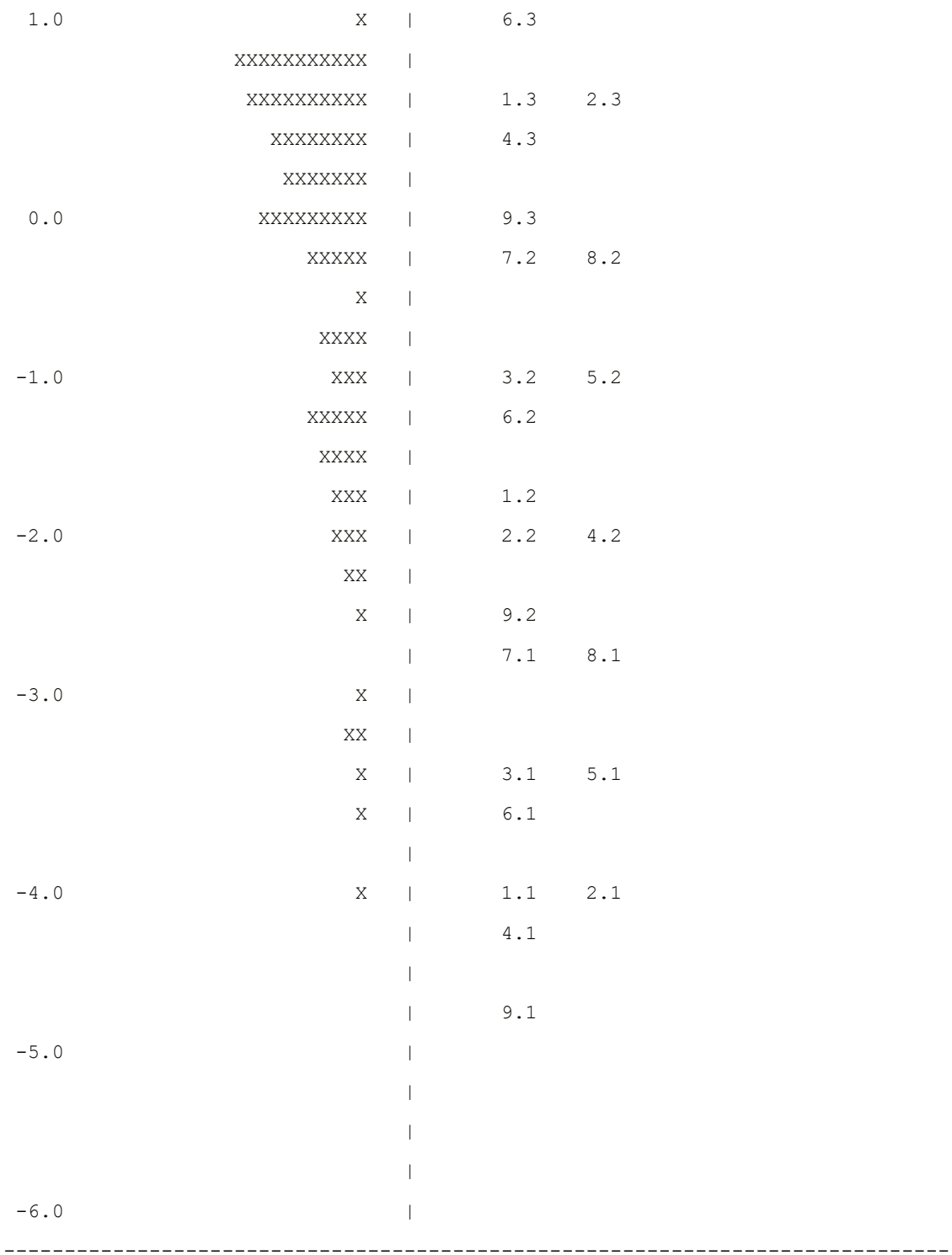
Wechsler, D. (2004). *The Wechsler intelligence scale for children* (4th ed.). London: Pearson Assessment.

Wright, B. (1998). Estimating measures for extreme scores. *Rasch Measurement Transactions, 12*(2), 632–633.

Wright, B., & Masters, G. (1982). *Rating scale analysis: Rasch measurement*. Chicago: MESA Press.

Appendix A: Item-person map (Wave 7)



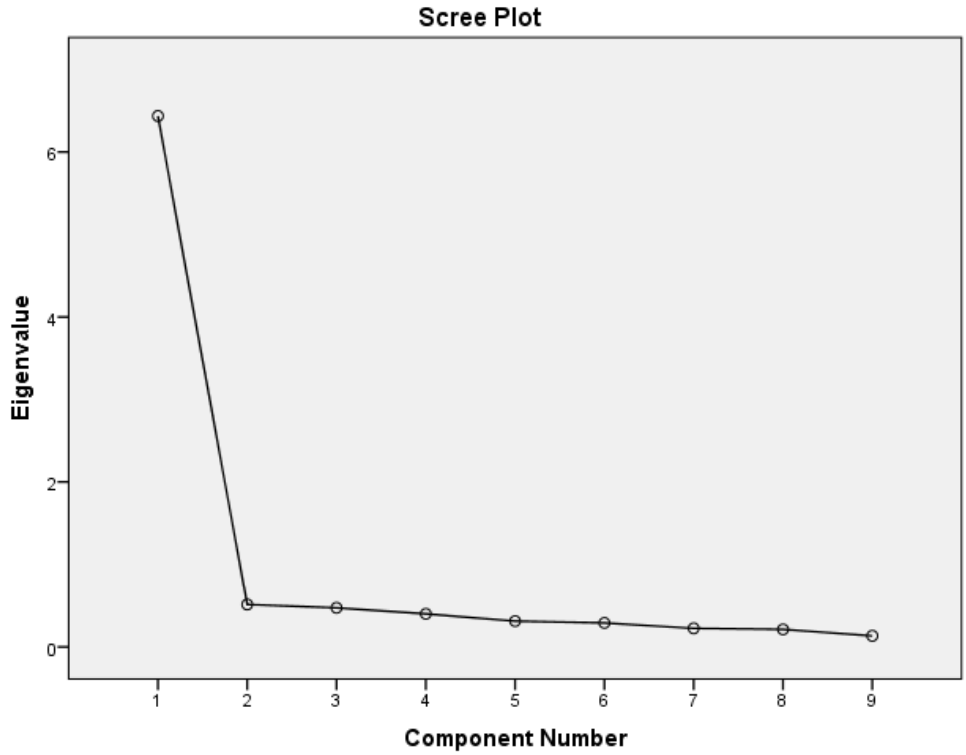

1.0 X | 6.3
XXXXXXXXXXX |
XXXXXXXXXX | 1.3 2.3
XXXXXXXX | 4.3
XXXXXXX |
0.0 XXXXXXXXX | 9.3
XXXXX | 7.2 8.2
X |
XXXX |
-1.0 XXX | 3.2 5.2
XXXXX | 6.2
XXXX |
XXX | 1.2
-2.0 XXX | 2.2 4.2
XX |
X | 9.2
| 7.1 8.1
-3.0 X |
XX |
X | 3.1 5.1
X | 6.1
|
-4.0 X | 1.1 2.1
| 4.1
|
| 9.1
-5.0 |
|
|
|
-6.0 |
Each X represents 7 students

Appendix B: Principal component analysis (Wave 7)

Total variance explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.437	71.517	71.517	6.437	71.517	71.517
2	.514	5.714	77.231			
3	.474	5.267	82.498			
4	.401	4.452	86.950			
5	.313	3.475	90.425			
6	.290	3.224	93.648			
7	.225	2.500	96.148			
8	.212	2.359	98.507			
9	.134	1.493	100.000			

Extraction Method: Principal Component Analysis.



Component Score Coefficient Matrix

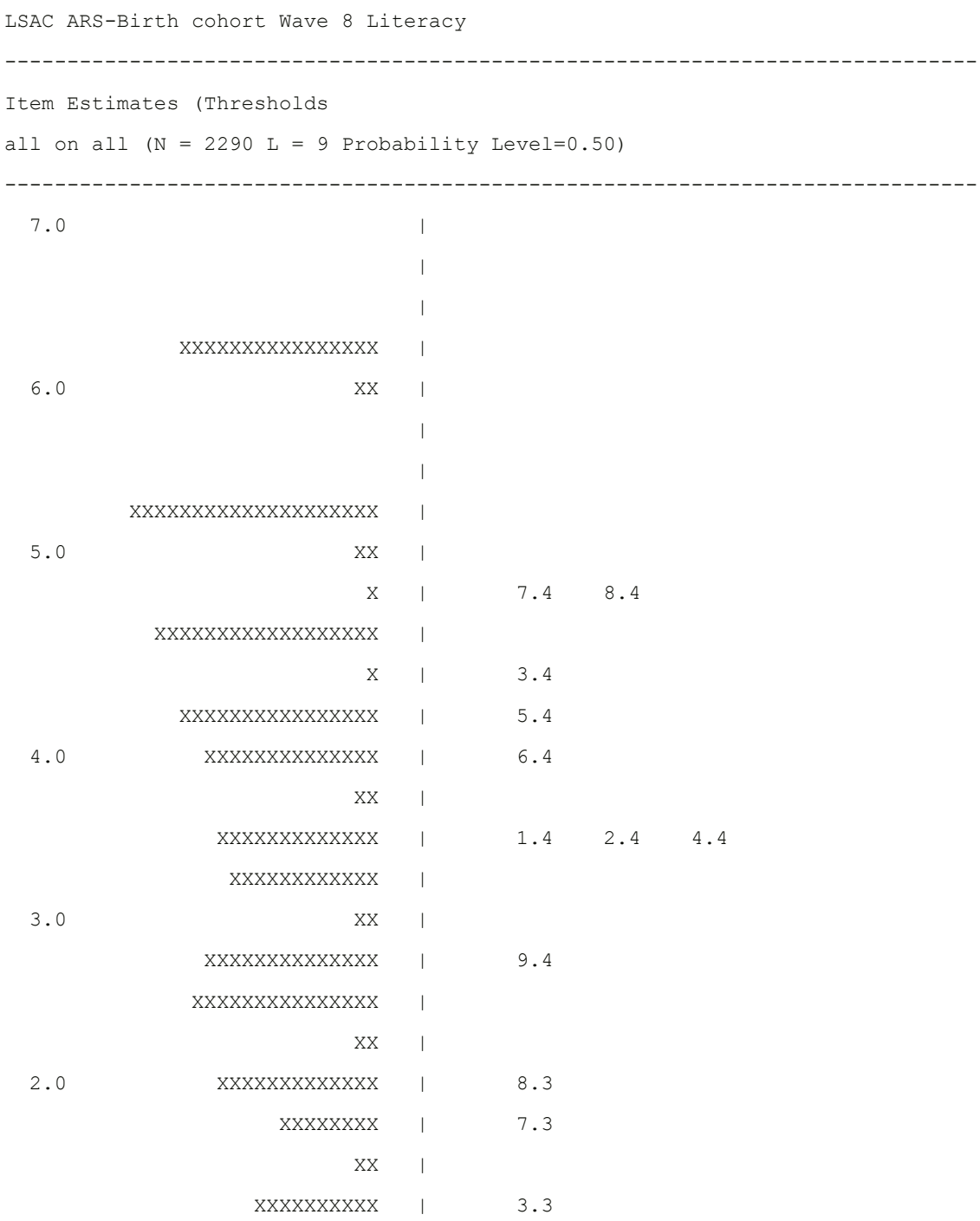
	Component 1
glc09a11 16.1 - Conveys ideas when speaking	.123
glc09a20 16.2 - Understands and interprets a story read aloud	.133
glc09a12 16.3 - Strategies to gain information	.132
glc09a13 16.4 - Reads fluently	.135
glc09a21 16.5 - Reads and comprehends expository text	.132
glc09a16 16.6 - Composes multi-paragraph texts	.137
glc09a17 16.7 - Redrafts writing	.138
glc09a18 16.8 - Makes editorial corrections	.134
glc09a19 16.9 - Uses computer for variety of purposes	.118

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalisation.

Acknowledgement

This appendix is prepared by Sam Rothman from the Australian Council for Educational Research.

Appendix C: Item-person map (Wave 8)



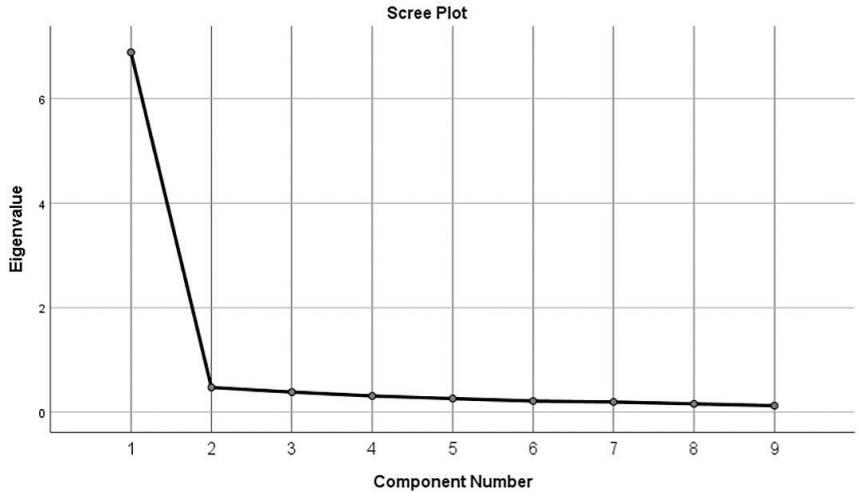
```
 1.0                XXXXXXXXXX   |        5.3
                    XXXXXXXXXX   |        6.3
                       XXXXXXX   |        1.3
                             X   |        2.3     4.3
                         XXXXX   |
 0.0                      XXXX   |
                       XXXXXXX   |        7.2     9.3
                          XXX    |        8.2
                         XXXXX   |
-1.0                      XXX    |        3.2
                          XXX    |        5.2
                          XXX    |        6.2
                           XX    |        1.2
-2.0                       XX    |        2.2     4.2
                            X    |        7.1     8.1
                            X    |
                            X    |        9.2
-3.0                        X    |        3.1
                            X    |        5.1
                            X    |        6.1
                                 |        1.1
                                 |        2.1     4.1
-4.0                        X    |
                                 |
                                 |        9.1
                                 |
-5.0                             |
                                 |
                                 |
                                 |
-6.0                             |
------------------------------------------------------------------------------
  Each X represents    7 students
==============================================================================
```

Appendix D: Principal component analysis (Wave 8)

Total variance explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.886	76.511	76.511	6.886	76.511	76.511
2	0.472	5.243	81.754			
3	0.383	4.257	86.011			
4	0.310	3.440	89.451			
5	0.260	2.892	92.342			
6	0.212	2.358	94.701			
7	0.194	2.159	96.860			
8	0.159	1.766	98.626			
9	0.124	1.374	100.000			

Extraction Method: Principal Component Analysis.



Component Score Coefficient Matrix

	Component 1
hlc09a11 17.1 – Conveys ideas when speaking	.121
hlc09a20 17.2 – Understands and interprets a story read aloud	.127
hlc09a12 17.3 – Strategies to gain information	.129
hlc09a13 17.4 – Reads fluently	.128
hlc09a21 17.5 – Reads and comprehends expository text	.132
hlc09a16 17.6 – Composes multi-paragraph texts	.130
hlc09a17 17.7 – Redrafts writing	.131
hlc09a18 17.8 – Makes editorial corrections	.127
hlc09a19 17.9 – Uses computer for variety of purposes	.117

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalisation.