Department for Education

Study of Early Education and Development (SEED): Wave 7

Technical report

June 2025

Svetlana Speight, Sehaj Bhatti, Anna Keyes, Rebecca Parker – the National Centre for Social Research

Government Social Research

Contents

List of tables

1. Overview

1.1. About the SEED study

The Study of Early Education and Development (SEED) is a major longitudinal study following nearly 6,000 children from across England from age 2. SEED so far has included seven waves of data collection:

- Face-to-face surveys of families and children in 2013-2018 when the children were 2, 3, 4 and 5 years old (waves 1-4)
- Two specially commissioned COVID-19 web-CATI (Computer Assisted Telephone Interviewing) surveys: in September-October 2020 (wave 5) and in May-June 2021 (wave 6), when children were aged 8-10 years
- A face-to-face survey of families and children in 2022-2023, when children were in year 6 at school and aged 10-11 years old (wave 7)

Information was collected from the child's parents or guardian at all seven waves, and cognitive assessments of children were administered at waves 2, 3, 4 and 7. In addition, teacher surveys were administered in wave 4 when children were in year 1 at school, and in wave 7 when children were in year 6. Information collected in the interviews is linked with information from the National Pupil Database (NPD) to track children's progress. This so far has included the Early Years Foundation Stage Profile, Phonics, Key Stage 1, and Key Stage 2 data.

The SEED study was originally conceived as an evaluation of a new early years policy, which offered two-year-olds from low-income families funded hours of early childhood education and care (ECEC). Two-year-old children living in the 20% lower income households, as well as those looked after by their local authority, became eligible for 15 hours of funded early education per week for 38 weeks of the year in September 2013. In September 2014, this was extended to children in the 40% lower income households in England, children with special needs and those who had left care. SEED was designed to cover families across the spectrum of eligibility so that comparisons could be made that explore the effect of eligibility on children's outcomes. The SEED study includes families whose children were born across six consecutive academic terms, covering two complete academic years.

The oldest children in SEED were born between September and December 2010 (cohort 1), and the youngest children were born between April and August 2012 (cohort 6). Children from the most disadvantaged (20%) families in cohort 1 were eligible for the two-year-olds offer for just one term, and then they became eligible for the three- and four-year-olds offer instead. Children from the moderately disadvantaged (20-40%) families in

cohort 1 were not eligible for the two-years-olds offer at all. In contrast, for cohort 6, children from both the most disadvantaged (20%) and moderately disadvantaged (20-40%) families were eligible for all three terms of the two-year-olds offer.

Table 1 shows when different cohorts of children in the study were in which school year during their primary school period.

Table 1: School years for SEED cohorts

	Older cohorts (Cohorts 1-3)	Younger cohorts (Cohorts 4-6)
Reception	2015-16	2016-17
Year 1	2016-17	2017-18
Year 2	2017-18	2018-19
Year 3	2018-19	2019-20
Year 4	2019-20	2020-21
Year 5	2020-21	2021-22
Year 6	2021-22	2022-23

The SEED study is funded by the Department for Education. Since 2021, it has been carried out by the National Centre for Social Research (NatCen) in collaboration with University College London, Durham University, University of Bristol and SQW. Between 2013-2021, the study was carried out by NatCen in collaboration with the University of Oxford, Frontier Economics and Action for Children.

The wave 7 of the longitudinal study received an approval from NatCen's Research Ethics Committee on 23rd September 2021, and a separate approval of the process for matching survey data to the National Pupil Database (NPD) on 12th July 2023.

2. Sampling

2.1. Sampling frame

The sampling frame for the longitudinal survey was Child Benefit records. This was considered to be an appropriate sampling frame because until January 2013 it was a universal benefit, with a take-up rate of around 98%. Although changes came into effect in January 2013 that affected Child Benefit records as a universal sampling frame,[1] HMRC estimated that at the time, 90% of families in the Child Benefit population would continue to receive some or all of their Child Benefit.[2] Furthermore, as the range of dates of birth determining eligibility to be selected for the SEED study was from September 2010 to August 2012, it was assumed that the changes to the Child Benefit would not have a substantially negative impact on the coverage of the eligible population in the Child Benefit records, and a spread of income groups would be adequately covered.

2.2. Clustering

Face-to-face surveys are often geographically clustered to improve fieldwork efficiency. For this study, clustering was particularly important because of the desire to assess the quality of early years and childcare settings used by parents. In many areas (particularly urban areas), a large number of settings are available locally, and without adequate clustering we would have found that many settings would have been used by just one family in our achieved sample. To improve the chance that families in our achieved sample used the same settings as each other, we used two stages of clustering for SEED. First, we selected postcode districts (or groups of postcode districts) as Primary Sampling Units (PSUs), followed by three postcode sectors (or groups of sectors) as Secondary Sampling Units (SSUs) within each PSU. This meant that the cost of assessing a setting's quality involved a more efficient use of money because the score could be associated with several children instead of just one.

In practice the sampling was done in three stages:

- 111 PSUs were selected in proportion to a weighted sum of the number of eligible families within each PSU (with weights calculated to reflect the final desired proportions of the three disadvantage groups, see below)

[1] In January 2013, the High Income Child Benefit Charge was introduced. It applied to anyone with an adjusted net income over £50,000 who received Child Benefit, or whose partner received it. This was a stepped charge, and families where either of the parents' income was over £60,000 became effectively ineligible to receive the Child Benefit.
[2] HMRC (2012) Child Benefit: Income Tax Charge on Those with Higher Incomes.

- Three SSUs were selected within each PSUs in proportion to a weighted sum of the number of eligible families within each SSU
- Five or six families in each disadvantage group were selected within each SSU in proportion to their weights.

2.3. Disadvantage groups

To maximise our ability to make comparisons of child outcomes across the spectrum of eligibility for funded early education for two-year-olds, each cohort of children within SEED was designed to have three subgroups:

(1) the 20% most disadvantaged families

(2) moderately disadvantaged families (20-40%)

(3) the 60% least disadvantaged families.

The three subgroups were sampled in equal proportion, i.e., such that each group made up around a third of the sample in each cohort. As the three groups were not of equal size in the population, a weighted sampling approach was used to create as close to an equal probability sample as possible. The selection weights were equal to the ratio of the desired proportion (one third) to the population proportion in each cohort.

Families were put into groups by Department for Work and Pensions (DWP) prior to sampling using the following criteria:

- The *20% most disadvantaged families* had a parent in receipt of one of the following benefits or tax credits: OBJ
    - Income-based Jobseeker's Allowance (JSA-IB)
    - Income-related Employment Support Allowance (ESA-IR)
    - Income Support (IS)
    - Guaranteed element of the State Pension Credit (PC with Guarantee Credit)
    - Child Tax Credit *only* (not in receipt of an accompanying Working Tax Credit award) with household gross earnings of less than £16,190.
- The *moderately disadvantaged group (20-40%)* had a parent in receipt of Working Tax Credits with household gross earnings of less than £16,190. OBJ
- The *60% least disadvantaged group* had parents not in receipt of any of the qualifying benefits or tax credits.

2.4. Sample for wave 7

For SEED wave 7, the eligible sample consisted of families from wave 3 of the longitudinal study who had agreed to be contacted for future research and had not withdrawn from the study since then. All those who completed wave 3 had also completed waves 1 and 2. Those participants who had not taken part in the SEED study since waves 1 or 2 were not invited to take part in wave 7. This decision was based on the estimated very low probability of reengaging these research participants with the study. This approach was in line with the approach taken to keep-in-touch communications with research participants between waves of data collection since 2018, where only those who had taken part in wave 3 were sent newsletters about the study.

Availability of a postal address was another condition for who could be included in the wave 7 issued sample, as without an address, it would not have been possible to issue the family to face-to-face fieldwork. A pre-notification campaign and a tracing exercise took place ahead of each fieldwork period (see section 4.2). Those respondents who had an address available at the end of those exercises were included in the issued sample.

3. Questionnaires

3.1. Overview

For wave 7, there were three parts to the interviewer visit:

1.) Parent questionnaire

2.) Child questionnaire

3.) Child cognitive assessments

The median length of the interviewer visit was 61 minutes. The questionnaire was developed by the National Centre for Social Research in collaboration with the Department for Education, University College London, Durham University, University of Bristol and SQW.

Consent was sought from the parent/carer for the parent questionnaire, and from both the parent/carer and the child for the child questionnaire and the cognitive assessments. Consent was recorded electronically on the interviewer's laptop.

The questionnaire was programmed in Blaise 4, a software suite which is used for face-to-face fieldwork.

3.2. Parent questionnaire

For the parent questionnaire, most of the questions were asked by the NatCen interviewer of the parents/carer. Where questions were self-completed (CASI), the parent/carer entered the answers directly into the computer programme. Table 2 outlines the content of the parent questionnaire.

Table 2: SEED wave 7 parent questionnaire topics

Section	Subtopics included
A. Eligibility	▪ Whether same respondent as main contact in SEED sample ▪ Eligibility
B. Health, SEN, and health behaviours	▪ Whether child has health condition or illness ▪ Child considered to have a special educational need (SEN) ▪ Education, Health and Care (EHC) plan ▪ Child participation in physical activities ▪ Child morning and bedtime routine
C. Activities	▪ Child involvement in activities outside of school hours ▪ How often child reads for enjoyment
D. Digital environment	▪ Child access to internet at home for schoolwork

	▪ Child general internet access and usage at home and outside of the household ▪ Child internet and social media profiles ▪ How many hours the child spends online on a school day
E. Social networks and relationships	▪ How often the child socialises with family and friends outside of the household
F. Parent-child relationship	▪ How often parent-child eat together ▪ Parent-child conversations ▪ Parent involvement with school
G. Parent's academic expectations for child	▪ GCSEs ▪ University
H. Child socio-emotional development **(self-completion)**	▪ Strengths and difficulties questionnaire (SDQ)
I. Parent mental health **(self-completion)**	▪ Kessler 6 scale
J. Parenting Styles **(self-completion)**	▪ Parenting Styles and Dimensions Questionnaire (PSDQ)
K. Household chaos and unpredictability	▪ Household chaos and unpredictability: CHAOS scale
L. Adverse events	▪ Employment ▪ Income ▪ Accommodation ▪ Serious illness in the household ▪ Family death
M. Family socio-demographics	▪ Employment ▪ Qualifications ▪ Income ▪ Tenure ▪ Whether child is supported by Social Care ▪ Eligibility for and receipt of free school meals
N. Administration	▪ Permission to contact class teacher and their contact details ▪ Update of contact information for respondent and their 'stable contact'

3.3. Child questionnaire

In addition to the parent/carer interview, there was also a 10-15-minute questionnaire for the child. The child completed it via audio Computer Assisted Self Interview (audio-CASI). Table 3 outlines the content of the child questionnaire.

Table 3: SEED wave 7 child questionnaire topics

Section	Subtopics included
A. Attitudes to school	▪ Learning ▪ Teachers ▪ School work ▪ Playtime ▪ Bullying
B. Relationships with friends and families	▪ How the child feels about their friends ▪ Time spent with friends ▪ How the child feels about their family ▪ How often parent-child conversations happen ▪ How often child speaks to another adult in their family
C. Self-esteem	▪ How the child feels about themselves (Rosenberg Self-esteem scale, 5 items)
D. Reading for enjoyment	▪ Reading frequency ▪ Reading enjoyment
E. Mental health	▪ How the child is feeling (Short Mood and Feelings Questionnaire, 13 items)
F. Digital environment	▪ Internet access and usage within the household ▪ Internet use and social media profiles ▪ Internet risks ▪ Cyberbullying

3.4. Cognitive assessments

In addition to the questionnaire for the child, the interviewer also administered the Listening Comprehension section of the Wechsler Individual Achievement Test (WIAT) III: Receptive Vocabulary and Oral Discourse Comprehension scale. [OBJ]

For the Receptive Vocabulary sub-test, children were shown four pictures on a single page and then asked to select the picture which contains the item correctly matching a word given by the interviewer. For the Oral Discourse Comprehension sub-test, children listened to a series of audio recordings (tracks) and were then asked one, or sometimes two, questions by the interviewer about what they had just heard. Questions were not multiple choice, instead children had to answer in their own words. The assessment

ended if the child had given the wrong answer (or did not know the answer) to four consecutive questions.

4. Survey Fieldwork

4.1. Piloting

The survey questionnaires and processes were tested in a pilot. The pilot sample was made up of 96 families in total. Of those, 57 respondents were from the original pilot panel, and 39 were main stage respondents being used for the pilot to boost numbers. All respondents had previously participated in waves 1, 2 and 3 of SEED, so the approach was the same as with the main stage sample (see section 2.4).

First, a pre-engagement letter and an email were sent to the sample eligible to take part in the pilot. The aim was to engage research participants ahead of fieldwork and to check that their contact details on file were up to date. Once that exercise had been completed, the pilot fieldwork started on 1st November 2021. The end date for the fieldwork was 21st November 2021 for the respondents from the original pilot panel and 12th December 2021 for the main stage panel respondents.

Five interviewers in total worked on the pilot. They were briefed by researchers on the project on MS Teams on 25th-26th October 2021 (x4), and on 8th November 2021 (x1). On 22nd November, all five interviewers attended a debrief session on MS Teams to provide feedback on the pilot. This was with the NatCen research team and the project manager from the Department for Education (DfE). Ahead of the de-brief session, each interviewer collated their feedback in a pilot feedback form, all of which was discussed during the session.

Interviews were completed with 63 SEED families in total. The response rate achieved was 66%. Table 4 shows a full breakdown of response rates.

Table 4: Wave 7 SEED CAPI pilot response rates

	Original pilot cases	Main stage cases	Total pilot sample
Issued	57	39	96
Productive (n)	38	25	63
Productive (%)	**67%**	**64%**	**66%**
Refusal (n)	11	9	20
Non-contact (n)	4	3	7
Other unproductive (n)	4	2	6
Ineligible (n)	0	0	0

Respondents who were from the main stage panel were eligible for an incentive if they belonged to the most disadvantaged sample group or were from the moderately disadvantaged group and had not taken part in either of the COVID follow-up waves (waves 5 and 6). 17 respondents were eligible in total, and of those, ten received an incentive in the form of a post office voucher for £20, which respondents could exchange for cash at a post office branch. It was sent to respondents by post with a thank you letter in week commencing 20th December 2021. Respondents from the pilot panel were not eligible for an incentive. (The approach to incentives during the main stage of wave 7 is outlined in section 4.5.)

Interviewers reported that making contact went smoothly and co-operation to take part was high. Most respondents still remembered the SEED study – some even recognised their interviewers – and respondents were mostly keen to participate again. The children generally did not remember taking part before, but some remembered, and still used, the water bottles given out during a previous wave. Concerns around COVID-19 from respondents were minimal overall and did not significantly impact fieldwork. Most respondents felt comfortable to participate and to invite the interviewer into their home. Respondents were happy to comply with NatCen's COVID-19 protocols such as opening windows and maintaining social distancing and interviewers reported no issues around this.

A few changes were implemented for the main stage as a result of interviewer feedback during piloting:

1) More information was added to the Address Record Form used by interviewers during their visits. Namely, a flag was added to indicate whether an (informal) interpreter was used on the last face-to-face wave, as well as information about when the family last participated in a face-to-face wave and whether they participated in the COVID web-CATI surveys (waves 5 and 6)
2) More instructions were added to interviewer materials to support them in discussing incentives with eligible participants on the doorstep.

The questionnaire in the pilot was mostly the same as the main stage questionnaire, as no major issues were discovered during the piloting. There were two main differences:

1) The cognitive assessment with children in the pilot included the Receptive Vocabulary scale only, as it had not been possible to program the Oral Discourse Comprehension (ODC) scale in time for the pilot fieldwork. The ODC assessment was added to the questionnaire for the main stage.
2) The section collecting details of the school attended by the children was redesigned to make it easier for interviewers to navigate and make the questionnaire less vulnerable to interviewers making mistakes.

The median length of the interviewer visit in the pilot was 49 minutes. Interviewers did not report any issues with the interview length, and this was true for both parents and children. Interviewers usually gave families a choice about which elements to complete first. Most families started with the parent interview and then went on to the child cognitive WIAT-III assessment and finally the child questionnaire. However, some children preferred to be first and interviewers were flexible and able to respond to families' needs and preferences.

Families from the main stage panel who took part in the pilot were included in the final analysis dataset, although children in this group missed data on the ODC scale.

4.2. Pre-notifications and tracing

Ahead of the main stage fieldwork, pre-notification and tracing activities took place (separately for Cohorts 1-3 in autumn 2021 and Cohorts 4-6 in autumn 2022), which aimed to check that the postal address in the sample database was not missing and was up to date. This involved contacting research participants by letter, email, and phone as necessary and offering to enter them into a prize draw to win an iPad if they confirmed or updated their address. Where an address was missing and it was not possible to get in touch with research participants themselves after a few attempts, NatCen's Telephone Unit tried to contact participants' 'stable contact' (a relative or friend). All address updates were implemented ahead of issuing the sample to survey fieldwork. Where a postal address was still missing after all pre-notification and tracing activities, those participants were not issued to wave 7 fieldwork.

4.3. Interviewer briefings

Interviewer briefings were delivered by the research team on MS Teams over two half-days for each group of interviewers. For Cohorts 1-3, briefings took place from March to July 2022, and 108 interviewers in total were briefed to work on the project. For Cohorts 4-6, whose fieldwork was taking place about a year later, interviewers who had worked on SEED with Cohorts 1-3 were asked to self-brief using self-briefing instructions including links to videos about cognitive assessments. Interviewers new to SEED were briefed in MS Teams briefings by the research team over two half-days for each group in the period from December 2022 to March 2023. In total, 62 interviewers were briefed to work on SEED in 2023 (not including self-briefing interviewers). At the end of each briefing, interviewers were assessed by a member of the research team on how well they could conduct cognitive assessments and were formally 'signed off' to work on SEED.

4.4. Fieldwork dates

Data collection took place when children were in year 6 at school. Cohort 1-3 families (1093 families) completed the survey from March to August 2022. Families in Cohorts 4-6 (1028 families) completed the survey in the following school year, from January to July 2023.

4.5. Incentives

A sub-sample of participants in the face-to-face survey was eligible for incentives. The eligibility was based on the following criteria:

- Respondent belongs to the most disadvantaged sample group OR
- Respondent belongs to the moderately disadvantaged sample group AND respondent did not take part in either of the COVID follow-ups (waves 5 and 6)

Incentives were slightly different in 2022 and 2023:

- 2022 fieldwork: a £20 Post Office voucher sent in a thank you letter (i.e., conditional)
- 2023 fieldwork: a book of eight 1st class postage stamps included in the advance letter (i.e., unconditional) AND a £15 Post Office voucher sent in a thank you letter (i.e., conditional)

There was no evidence that changing the approach to incentives between 2022 and 2023 had an impact on response from the eligible group.

4.6. Response rates

The overall response rate was 61%, with a slightly higher response in 2023 (62%) compared with 2022 (60%), and the highest response from the main stage respondents in the pilot (64%,Table 5).

Table 5: Fieldwork response figures, by fieldwork period

Outcome	Total	2021 (Cohort 1-3 cases in pilot)	2022 (Cohorts 1-3)	2023 (Cohorts 4-6)
Issued	3542	39	1822	1681
	100%	100%	100%	100%
Ineligible	20	0	7	13
	0.6%	0%	0.4%	0.8%
Eligible	3522	39	1815	1668
	100%	100%	100%	100%
Productive	2146	25	1093	1028
	61%	**64%**	**60%**	**62%**
Refusal	771	9	392	370
	22%	23%	22%	22%
Other unproductive (ill, away, language difficulties)	57	2	36	19
	2%	5%	2%	1%
Non-contact	548	3	294	251
	16%	8%	16%	15%

The response was the lowest among the most disadvantaged families (49%), higher among the moderately disadvantaged (59%) and the highest among the least disadvantaged (70%; see Table 6).

Table 6: Fieldwork response figures, by disadvantage group

Outcome	Total	Most disadvantaged	Moderately disadvantaged	Least disadvantaged
Issued	3542	851	1242	1449
	100%	100%	100%	100%
Ineligible	20	4	8	8
	0.6%	0.5%	0.6%	0.6%
Eligible	3522	847	1234	1441
	100%	100%	100%	100%
Productive	2146	416	728	1002
	61%	**49%**	**59%**	**70%**
Refusal	771	192	305	274
	22%	23%	25%	19%
Other unproductive (ill, away, language difficulties)	57	16	25	16
	2%	2%	2%	1%
Non-contact	548	223	176	149
	16%	26%	14%	10%

5. Data processing

Once data collection had been completed, participants' free text responses under "other" answer options were coded up and incorporated into the original responses. The dataset was checked for errors, any inconsistencies were investigated and corrected. Variable labels were checked and improved. A number of derived variables (e.g., scales derived for multi-item measures) were added to the dataset. The following geography variables were also added based on participants' postal address at the time of the survey:

- Index of Multiple Deprivation (IMD), England, 2019 score, deciles
- Income Deprivation Affecting Children Index (IDACI), deciles
- Region
- 2011 census rural-urban classification

Participants' names, contact details and dates of birth as well as other potentially disclosive data were removed from the analysis dataset.

6. Weighting

6.1. Overview

All respondents who took part in SEED waves 1, 2 and 3, and who had not withdrawn from the study since then and had not become ineligible, were issued at wave 7. The exception was respondents for whom we did not have a postal address, who could not be issued.

Three wave 7 weights were produced:

1. a cross-sectional weight for all cases that responded in waves 1 to 3 and wave 7,
2. a longitudinal weight for all cases that responded in every wave from 1 to 7,
3. a partial longitudinal weight for all cases that responded in waves 1 to 4 and wave 7.

The cross-sectional weight should be used for analysis of the SEED wave 7 data in its own right (e.g., prevalence of behaviours and attitudes in the wave 7 data). The cross-sectional sample consists of 2146 cases that responded in wave 7 and previously completed waves 1 to 3.

The longitudinal weight should be used for longitudinal analysis of cases with complete data for all seven waves of SEED. The longitudinal sample consists of 1109 cases that responded in wave 7 and previously completed waves 1 to 6.

The partial longitudinal weights should be used for analysis of cases that completed the face-to-face waves of SEED. The partial longitudinal sample consists of 1984 cases that responded in wave 7 and previously completed waves 1 to 4. This sample disregards the 'Covid waves' 5 and 6.

6.2. Cross-sectional weight

Non-response at wave 7 was modelled using logistic regression with a dichotomous outcome variable (1=response; 0=non-response). The model was weighted by wave 3 non-response weight and only those families assumed to be eligible were included. Measures from both the wave 1 and wave 3 surveys were tested as predictors in the model. This is because some questions were asked again after wave 1 but others were not.

The following variables (taken from wave 1 where not specified) were tested as predictors of response:

- Cohort * Disadvantage group (18 categories indicating the cohort 1-6 and disadvantage group);
- Sex of child

- Ethnicity of child (White/BAME)
- Age group of parent (grouped)
- Number of adults in the household (at wave 3)
- Number of children under 15 in the household (at wave 3)
- Number of siblings to the sample child (at wave 3)
- Number of birth parents to the sample child (at wave 3)
- Whether respondent living with spouse/partner (at wave 3)
- Whether couple or lone parent household (at wave 3)
- Whether working household or not (at wave 3)
- Whether mother works or not (at wave 3)
- Mother's level of qualifications
- Household work status (at wave 3)
- Whether claiming any benefits
- Household income (grouped) (at wave 3)
- Tenure (at wave 3)
- Whether English a first language
- Whether speak English with child (mainly)
- Region (based on wave 7 postcode)
- IMD quintile (based on wave 7 postcode)
- Population density quintile (based on wave 7 postcode)
- Urban/rural indicator (six groups) (based on wave 7 postcode)

The variable indicating cohort/disadvantage group was fixed in the model along with region and sex of child. This ensured that bias was minimised for these measures, regardless of whether they were significant predictors of response. A forward stepwise logistic regression was used to select the other predictors and double-checked using backwards stepwise logistic regression, which produced a similar model. The final model included the following variables:

- Cohort * Disadvantage group (18 categories indicating the cohort 1-6 and disadvantage group);
- Region
- Sex of child
- Age group of parent (grouped)
- Whether a couple or lone parent household (at wave 3)
- Mother's level of qualifications
- Tenure (at wave 3)
- Whether speak English with child (mainly)
- Population density quintile (based on wave 7 postcode)
- IMD quintile (based on wave 7 postcode)

Non-response weights were created as the inverse of the predicted probability of response. The top 0.5% of these non-response weights was trimmed back to the 99.5th percentile. These weights were then multiplied by the wave 3 weights to produce a final cross-sectional weight for wave 7. The top weight was trimmed, then the weights were scaled to have a mean value of 1.

6.3. Longitudinal weight

Non-response at wave 7 was modelled using logistic regression with a dichotomous outcome variable (1=response; 0=non-response). The model was weighted by wave 6 longitudinal weight and only those families assumed to be eligible were included. This is a similar approach to that used for the cross-sectional weights; for the longitudinal weights, measures from both the wave 1 and wave 4 surveys were tested as predictors in the model.

The following variables (taken from wave 1 where not specified) were tested as predictors of response:

- Cohort * Disadvantage group (18 categories indicating the cohort 1-6 and disadvantage group);
- Sex of child
- Ethnicity of child (White/BAME)
- Age group of parent (grouped)
- Number of adults in the household (at wave 4)
- Number of children under 15 in the household (at wave 4)
- Number of siblings to the sample child (at wave 4)
- Number of birth parents to the sample child (at wave 4)
- Whether respondent living with spouse/partner (at wave 4)
- Whether couple or lone parent household (at wave 4)
- Whether working household or not (at wave 4)
- Whether mother works or not (at wave 4)
- Mother's level of qualifications
- Household work status (at wave 4)
- Whether claiming any benefits
- Household income (grouped) (at wave 4)
- Tenure (at wave 4)
- Whether English a first language
- Whether speak English with child (mainly)
- Region (based on wave 7 postcode)
- IMD quintile (based on wave 7 postcode)
- Population density quintile (based on wave 7 postcode)
- Urban/rural indicator (six groups) (based on wave 7 postcode)

As was the case for the cross-sectional weights, the variable indicating cohort/disadvantage group was fixed in the model along with region and sex of child. A forward stepwise logistic regression was used to select the other predictors then double-checked using backwards stepwise logistic regression, which produced an identical model. The final model included the following variables:

- Cohort * Disadvantage group (18 categories indicating the cohort 1-6 and disadvantage group);
- Region
- Sex of child
- Age group of parent (grouped)
- Number of siblings to the sample child (at wave 4)
- Whether working household or not (at wave 4)
- Tenure (at wave 4)
- Household income (grouped) (at wave 4)
- Population density quintile (based on wave 7 postcode)

Non-response weights were created as the inverse of the predicted probability of response. The top 0.5% of these non-response weights was trimmed back to the 99.5th percentile. These weights were then multiplied by the wave 6 longitudinal weights to produce a final longitudinal weight for wave 7. This had the highest three outlying weights trimmed, then was scaled to have a mean value of 1.

6.4. Partial longitudinal weight

Non-response at wave 7 was modelled using logistic regression with a dichotomous outcome variable (1=response; 0=non-response). The model was weighted by wave 4 weights and only those families assumed to be eligible were included. This is a similar approach to that used for the longitudinal weights. For the partial longitudinal weights, measures from both the wave 1 and wave 4 surveys were tested as predictors in the model.

The following variables (taken from wave 1 where not specified) were tested as predictors of response:

- Cohort * Disadvantage group (18 categories indicating the cohort 1-6 and disadvantage group);
- Sex of child
- Ethnicity of child (White/BAME)
- Age group of parent (grouped)
- Number of adults in the household (at wave 4)
- Number of children under 15 in the household (at wave 4)

- Number of siblings to the sample child (at wave 4)
- Number of birth parents to the sample child (at wave 4)
- Whether respondent living with spouse/partner (at wave 4)
- Whether couple or lone parent household (at wave 4)
- Whether working household or not (at wave 4)
- Whether mother works or not (at wave 4)
- Mother's level of qualifications
- Household work status (at wave 4)
- Whether claiming any benefits
- Household income (grouped) (at wave 4)
- Tenure (at wave 4)
- Whether English a first language
- Whether speak English with child (mainly)
- Region (based on wave 7 postcode)
- IMD quintile (based on wave 7 postcode)
- Population density quintile (based on wave 7 postcode)
- Urban/rural indicator (six groups) (based on wave 7 postcode)

As was the case for the other two sets of weights, the variable indicating cohort/disadvantage group was fixed in the model along with region and sex of child. A forward stepwise logistic regression was used to select the other predictors and double-checked using backwards stepwise logistic regression, which produced an identical model. The final model included the following variables:

- Cohort * Disadvantage group (18 categories indicating the cohort 1-6 and disadvantage group);
- Region
- Sex of child
- Age group of parent (grouped)
- Number of children under 15 in the household (at wave 4)
- Mother's level of qualifications
- Tenure (at wave 4)
- Whether working household or not (at wave 4)
- IMD quintile (based on wave 7 postcode)

Non-response weights were created as the inverse of the predicted probability of response. The top 0.5% of these non-response weights was trimmed back to the 99.5th percentile. These weights were then multiplied by the wave 4 weights to produce a partial longitudinal weight for wave 7. This was checked for outliers and left untrimmed, then scaled to have a mean value of 1.

7. Teacher survey

7.1. Sample

Out of 2146 children in the study (including those main stage panel participants who took part in the pilot, see section 7.3), 2115 attended school, 29 were educated at home, and there was missing data on school enrolment for two children. Where children were attending school, parents were asked for consent to contact the child's class teacher. Consent was given for 97% of children (2050 children), and the teacher's contact details were recorded for administration of the teacher survey. The teacher survey was conducted as a PAPI (pen and paper personal interview) survey, where questionnaires were sent to teachers by post.

7.2. Questionnaire

The questionnaire was four pages long (four sides of A4). It included questions about:

- The child's attitudes and behaviours in school
- The child's social-emotional development (Strengths and Difficulties Questionnaire, SDQ)
- The child's progress at school
- The child's experiences of bullying
- Parents' engagement with their child's schooling

Some questions in the questionnaire were the same as those asked of parents/carers (e.g., SDQ), whereas other questions were unique to the teacher questionnaire.

Teachers were asked to give answers on the basis of their knowledge of the child. If the child was no longer at their school or in their class, teachers were asked to answer on the basis of their knowledge of the child when they last attended their class or school.

There were no incentives for teachers taking part in the teacher survey.

7.3. Piloting

The pilot of the teacher survey aimed to test the methodology of the teacher survey on SEED Wave 7, including the questionnaire, communication strategy and the fieldwork approach. It was conducted with class teachers between 31st January and 1st April 2022, when children in the pilot were in year 6.

The pilot sample of teachers was obtained from the SEED Wave 7 pilot survey with families. In this pilot survey, parents/carers were asked to complete a questionnaire

themselves, and they were also asked for their consent for NatCen to contact their child's class teacher about completing the teacher survey questionnaire. A 100 per cent consent rate was achieved, with all 63 parents who had completed the parent questionnaire agreeing that their child's class teacher could be contacted. The parents were asked whether their child was still attending the same school as in year 1 (where those school details were available from SEED Wave 4) and if not, the details of their new school were collected. All parents were asked the name of their child's class teacher.

The teacher survey was sent to teachers of 58 children (of these, 22 children were from the pilot panel, and 36 children were from the main stage panel). We sent the survey out to fewer than 63 teachers because of the poor quality of school contact details collected from some families. On the basis of this, we changed the approach to collecting school details between the parent survey pilot and the main stage to obtain higher quality school contact information at the main stage. This was done by asking interviewers to look up details of all schools in a database, check them on screen and amend if necessary. We also made improvements to the school database that was used as a look-up file. The 58 children in the teacher survey pilot were taught by 50 different teachers across 39 different schools. This is because the sample included some teachers who taught more than one child taking part in the SEED study.

The questionnaires were administered as a postal survey. First, an advance letter was sent out to the headteachers followed by letters and questionnaires sent out to the teachers a week later. The fieldwork included two rounds of 'chaser' telephone calls from NatCen's Telephone Unit. Once the first round of chaser calls was complete, replacement questionnaires were sent to those teachers who had not yet completed the survey. The second round of chaser calls was after the teachers had received replacement questionnaires. This approach worked well and was followed for the main stage teacher survey as well.

There were three interviewers in total working on the phone calls, and two of them attended the de-brief along with the Telephone Unit (TU) Manager and Deputy Manager. Overall, interviewers fed back favourably on teachers' engagement with the SEED survey and their willingness to complete the questionnaires, especially in scenarios where they were already aware of SEED.

A total of 47 questionnaires were completed in the Wave 7 teacher survey pilot. This equates to 81% response rate (47 out of 58). At teacher level, the response rate was 80% (40 out of 50 teachers returned completed questionnaires). Forty teachers completed 47 questionnaires as there were some teachers who taught more than one child in the sample.

Analysis of pilot data showed that the proportion of missing data at different questions was very low – from 0% (i.e., the question was answered by all survey participants) to

2% (i.e., the question was answered by 98% of survey participants, while 2% left it blank). The distributions of answers looked as expected.

7.4. Fieldwork

Fieldwork for the main stage of the teacher survey was split into two main phases:

- September-December 2022 – fieldwork with teachers teaching children in Cohorts 1-3
- June-July & September-December 2023 – fieldwork with teachers teaching children in Cohorts 4-6

Communications included an advance letter to the headteacher to inform them about the study and to encourage them to support their staff with completing the questionnaires. This was followed a week later by a letter to the teacher, which included child questionnaire(s). If there were more than one child in the sample file who were taught by the same teacher, the teacher would receive questionnaires for each child.

There were two rounds of chaser phone calls to teachers. Round 1 phone calls focused on encouraging teachers to take part and on finding out if they had any queries or problems with completing the questionnaires. Questionnaire re-sends were mailed out after the first round of chaser calls was complete. Round 2 phone calls focused on the non-responding sample of teachers and aimed to encourage those slow or reluctant to take part to complete and send the questionnaires to NatCen by the deadline.

7.5. Response rates

The response rate achieved on the teacher survey across all phases was 77% (Table 7). Response was higher when the teacher survey took place when the child was still at their primary school in year 6, and it was lower if the survey was administered after the child had left their primary school.

Teacher survey data is available for 73% of children whose families completed the face-to-face survey, or for 1564 out of 2146 children.

Table 7: Teacher fieldwork response figures, by fieldwork period

Outcome	Total	2021 (Cohort 1-3 cases in dress rehearsal)	2022 (Cohorts 1-3)	2023 (Cohorts 4-6)
Parent completed the CAPI interview	2146	25	1093	1028
Parent completed the CAPI interview	100%	100%	100%	100%
Eligible for teacher survey: child enrolled at a school	2115	25	1072	1018
Ineligible: not enrolled or missing data	31	0	21	10
Parent consented to teacher survey	2050	25	1035	990
Issued to teacher survey	2038	22	1034	982
Issued to teacher survey	100%	100%	100%	100%
Teacher questionnaire completed	1564	18	756	790
Of all issued to teacher survey	**77%**	**82%**	**73%**	**80%**
Of all productive to CAPI survey	**73%**	**72%**	**69%**	**77%**

Note: 12 cases were not issued to the teacher survey because of insufficient contact information.

7.6. Data processing

Questionnaire data was keyed in house at NatCen using a Blaise 4 instrument. Most questionnaires (87%) were double-keyed (i.e., data was entered twice) for quality control, and any discrepancies were investigated and resolved.

7.7. Weighting

There are no separate teacher weights for analysis of teacher survey data. Once teacher survey fieldwork had been completed, tests were run to determine whether a separate set of weights for teacher responses would be required. Forward and backward stepwise

regression was used to fit a model of teacher response for productive cases eligible for the teacher questionnaire (n=2115). The variables tested for association with teacher response were child outcome measures, including Oral Discourse Comprehension scores, banded scores for emotional problems and other behaviours, scales for these behaviours, and scores for the self-esteem and moods and feelings questions. Regressions were run weighted by the cross-sectional weights and the demographic variables included in the cross-sectional non-response model forced in. Of the twenty child outcome variables tested, seven were found to have a significant ($p < 0.05$) association with teacher response in one or more regression models. Test weights for teacher non-response were created using a final model that included significant child outcomes and significant demographic variables from the cross-sectional model. These were then multiplied by the cross-sectional weights. The efficiency and residual bias for these combined test weights were compared with the existing cross-sectional weights and found to be extremely similar. The associations between teacher response and child outcomes were weak even when significant in regression modelling, therefore the test weights differed only marginally from the cross-sectional weights. We therefore concluded that separate teacher weights were not required for analysis of teacher responses.

8. Linking survey data with the National Pupil Database

8.1. Sample

As part of wave 7, SEED survey data collected from parents/carers, children and teachers was linked with Key Stage 2 data from the National Pupil Database (NPD). A sample file from the SEED study was created for linkage. This included all participants whose parents gave consent to link their child's data with NPD and who had not withdrawn from the study since then. The following types of data were included for each child where available: title, first name, surname, date of birth, sex, home postcode (most recent on record), academic year when the child was in year 6 at school, and NatCen ID. Once the survey data had been linked with NPD data, this personal data was removed from the linked file, and the linked file was deposited in the Office for National Statistics Secure Research Service (ONS SRS).

8.2. NPD variables

Variables requested from NPD included Key Stage 2 attainment results, Free School Meals status, Special Educational Needs status, absence and exclusions from school, Children in Need status and Children Looked After status.

8.3. Linked file

NPD data was successfully linked for 3349 participants in wave 3, out of 3930 whose survey data was available (85%). This linked file formed the basis of the main analyses in the key stage 2 impact report. Of the 2146 participants in wave 7, linked NPD data was available for 2111 participants (98%).

Department for Education

Reference: RR1528

ISBN: 978-1-83870-652-4

For any enquiries regarding this publication, contact us at: www.education.gov.uk/contactus

This document is available for download at www.gov.uk/government/publications