Evaluation of the Voice 21 Oracy Schools Programme

Natasha Plaister

Dave Thomson

January 2024

1. Executive summary

1.1 Main findings

- This report evaluates the effect of the Voice 21 Oracy Schools membership programme on attainment at Key Stage 2 and in the Early Learning Goals relating to communication and language at the end of the Early Years Foundation Stage in the 2021/22 academic year.
- It also looks at the impact on participants broken down into subgroups based on their level of engagement with the programme, and the length of time that they took part.
- The evaluation includes primary schools that joined the project between its launch in September 2020 and January 2022.
- This report did not find conclusive evidence that the programme had an impact on either Key Stage 2 attainment or on the likelihood of pupils achieving the relevant Early Learning Goals. However, while estimates of the impact were not statistically significant, they were positive for all of the outcomes measured.
- We did not find conclusive evidence to suggest that the programme had a higher impact on those schools that had taken part over a longer period.
- We found some evidence to suggest that the programme has more of an impact on KS2 reading in schools that reported higher levels of engagement from their teachers, although this didn't appear to be the case for the other outcomes.

1.2 Methodology

- This evaluation follows a quasi-experimental design. We used data from the National Pupil Database (NPD) to create a matched comparison group, similar to those schools who participated in the programme with respect to a set of variables.
- Participants were matched to non-participants using on nearest neighbour matching based on propensity scores.
- We then used regression models to compare the outcomes of pupils who attended schools in the matched comparison group to pupils in participating schools, using NPD data.

1.3 Limitations

- This evaluation uses a quasi-experimental design, which relies on creating a matched comparison group based on data from the NPD. This means that we are unable to control for factors not recorded in the NPD, for example levels of motivation.
- Of particular importance for this evaluation, the NPD does not include data on prior attainment for pupils working towards Early Learning Goals.
- Some comparison schools may have taken part in similar projects or received similar support from elsewhere. If this improved outcomes in comparison schools, it may have led to underestimation of effects.
- We are also limited to outcomes that are recorded in the NPD. The NPD does not include any direct measures of oracy skills, which would be much more closely related to the intervention than the outcomes actually used. Therefore, we are unable to measure the programme's impact on oracy skills directly.
- The programme has been running for a relatively short time and takes a whole school approach, largely working with school staff rather than directly with pupils.

It may be the case that the programme will have an impact on the outcomes measured, but that this will not become apparent until schools have participated for a number of years. If this is the case, effects would not yet be measurable as the programme has only been running since September 2020.

- This evaluation covers the period of the onset of the COVID-19 pandemic. This may have affected both the delivery and impact of the programme.

2. Introduction

Voice 21 works in partnership with teachers and schools across the UK with the aim of ensuring every child receives a high-quality oracy education.

In this evaluation, we looked at the impact of the programme on four outcomes: likelihood of achieving the Listening, Attention and Understanding (LAU) Early Learning Goal, likelihood of achieving the Speaking Early Learning Goal (both at the end of the Early Years Foundation Stage), reading attainment at Key Stage 2, and maths attainment at Key Stage 2. We include schools that joined the programme between its launch in September 2020 and January 2022, and look at outcomes in the 2021/22 academic year.

This work contains statistical data from ONS which is Crown Copyright. The use of ONS statistical data in this output does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data, The work uses research datasets which may not exactly reproduce National Statistics aggregates.

2.1 Methodology

This evaluation uses a quasi-experimental design. This involves comparing the outcomes of programme participants to those of pupils in a matched comparison group of schools that are statistically similar. This approach tries to mimic what would be done in a formal experiment such as a randomised control trial.

We used 1:1 nearest neighbour matching based on propensity scores. Some participating schools were not included in the assessment of Early Learning Goals, for example junior schools with no younger pupils, and others were not included in the assessment of KS2 attainment, for example infant schools with no older pupils. We therefore created two matched comparison groups: one for those participating schools that were included in the evaluation of Early Learning Goals, and one for those that were included in the evaluation of KS2 attainment. Some schools are included in the evaluation of both sets of outcomes.

Schools in the matched comparison groups are similar to participating schools with respect to the following matching variables:

- Region
- Phase (infant / junior / all)
- % of pupils eligible for FSM6
- % of EAL pupils
- Average KS1 prior attainment (for those to be assessed at KS2, where available)
- Historic (2019 and earlier) EYFS / Key Stage 2 outcomes, as appropriate

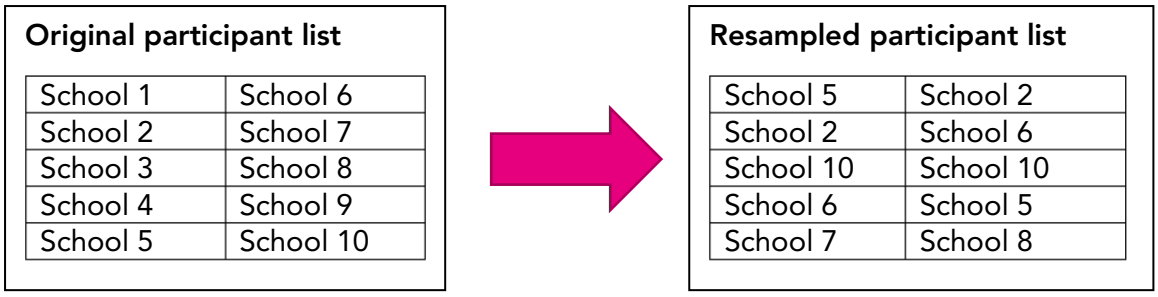
We then used regression models to compare outcomes for pupils in the participating schools to pupils in schools in the relevant matched comparison group. We controlled for pupils' prior attainment (for KS2 outcomes only), gender, ethnicity (ETC CHECK).

Confidence intervals were estimated using bootstrapping. While it is possible to construct confidence intervals simply by using the standard errors estimated by the regression models, this method only accounts for the uncertainty around the estimate made by the regression model; it does not account for the uncertainty in the matching process. Therefore, confidence intervals created in this way are likely to underestimate the standard errors and produce artificially narrow confidence intervals.

Bootstrapping allows us to take account of both sources of uncertainty. It involves repeatedly creating a new dataset by taking a random sample of participants from the original list, with replacement, then repeating the analysis using the fresh data. The random sample size will be the same as the size of the original list; if there were 100 participants in a given year, the random sample would also include 100 schools, although some participants would be included in the resampled list more than once, and some not at all.

The figure below shows an example of a resampled participant list, drawn from an original list of ten participants.

Figure 1: Resampling example



We repeat the process of creating and analysing new datasets 1,000 times. Our point estimates are found by taking the average of these 1,000 estimates, and the 95% confidence intervals are simply the range in which 95% of the 1,000 estimates lie.

We present estimates of the impact of the programme on outcomes overall, as well as outcomes broken down by the level of engagement with the programme. This includes estimates for length of participation (six months / one year / one and a half years / two years), level of engagement (low/mid/high, determined by proportion of teachers reported to be engaged with the programme).

For outcomes related to progression to university, we present estimates of the impact on the programme overall and broken down by gender. This is because progression to selective universities is a relatively rare event, and the sample size is not large enough to create reliable estimates of the impact on smaller subgroups.

We will also present estimates obtaining from the use of an alternative matching method as an appendix.

2.2 Data

Voice 21 provided a dataset consisting of information on all schools that participated in the project between September 2020 and January 2022. This included school identifiers (name, postcode, URN) and information on their participation in the project, including the length of participation and their responses to an evaluation and monitoring survey sent out by Voice 21. This dataset was linked to corresponding records in the National Pupil Database and publically available schools data.

The National Pupil Database is an administrative dataset maintained by the Department for Education, and includes records of achievements in national tests and examinations for all pupils who have been in state-funded education since 2002. For this evaluation, we used

data on attainment in Early Learning Goals and Key Stages 1 and 2, as well as some demographic variables.

The original dataset supplied by Voice 21 consisted of 468 schools. Some of these schools do not teach reception age children (e.g. junior and middle schools), and therefore are not included in the evaluation of Early Learning Goals. Others do not teach Year 6 children, and therefore are not included in the evaluation of KS2 attainment. Schools with both Reception and Year 6 pupils are included in the evaluation of both outcomes.

A number of schools could not be matched to records in the NPD, or did not have any attainment data available for the relevant year. These schools have been excluded from this analysis. Among others, this includes schools based in Wales, as the NPD only covers schools in England.

The final analysis included 405 schools for the evaluation of Early Learning Goals, and 395 schools for the evaluation of KS2 attainment.

Table 1: Number of schools included in analysis, by outcome

	Early Learning Goals	Key Stage 2
This outcome only	39	29
Both outcomes	366	366
TOTAL	405	395

3. Summary statistics and matching

This section begins with some summary statistics about participating schools. It will go on to describe the matching process used and how successful it was in creating a group of similar schools for comparison purposes.

3.1 Summary statistics

In this section, we look at some summary statistics describing the participating schools.

Voice 21 schools tend to have relatively high proportions of disadvantaged pupils, and the average IDACI score of participating schools was well above average. Voice 21 schools tend to have a relatively high proportion of pupils who have English as an additional language, and a low proportion of White British pupils, compared to schools nationally.

Table 2: Pupil characteristics of Voice 21 schools

	Voice 21 schools	National
% EAL pupils	28%	16%
% FSM6 pupils	32%	26%
% White British pupils	56%	71%
Mean IDACI score	0.22	0.17

In terms of attainment, Voice 21 schools tended to have slightly lower levels of attainment both in relevant Early Learning Goals and Key Stage 2 attainment prior to joining the programme, compared to schools nationally.

Table 3: Average prior attainment of schools before joining the programme

	Voice 21 schools	National
% achieving ELY in LAU	80%	82%
% achieving ELY in speaking	80%	82%
Mean KS2 maths score	102.9	102.9
Mean KS2 reading score	103.6	103.8

We broke participating schools down by length of participation and by level of engagement in the programme.

The schools included in this evaluation included four cohorts: those that joined in September 2020, March 2021, September 2021 and March 2022. The table below shows the number in each cohort that were included in this analysis.

Table 3: Number of schools included in analysis, by outcome and length of participation

Join date	Length of participation (years)	Early Learning Goals	Key Stage 2
Sept 2020	0.5	84	83
March 2021	1	172	162
Sept 2021	1.5	72	76
March 2022	2	77	74
TOTAL		405	395

Level of engagement was determined based on responses to a feedback survey administered by Voice 21 to teachers in participating schools, specifically by a question on whether teachers were actively applying the guidance received from Voice 21 in their teaching. Schools in which less than 33% of teachers were doing so were defined as low dosage, between 33 and 66% as mid dosage, and 66% or over as high dosage. Unfortunately, relatively few participating schools responded to the feedback survey; those that did not are excluded from the analysis broken down by level of engagement.

Table 4: Number of schools included in analysis, by outcome and level of engagement

% teachers actively applying learning	Level of engagement	Early Learning Goals	Key Stage 2
0-33%	Low	12	14
33-66%	Mid	69	68
66%+	High	49	49
No data	Unknown	275	264
TOTAL		405	395

3.2 Extent of success in creating matched comparisons

The matching process is intended to create a group of non-participants who are similar to the participating schools with respect to a set of matching variables, including prior attainment, pupil demographics and school characteristics such as region and phase of education. Any differences in the outcomes of this comparison group and the participating schools can then be assumed to be due to the programme.

Some participating schools were not included in the assessment of Early Learning Goals, for example junior schools with no younger pupils, and others were not included in the assessment of KS2 attainment, for example infant schools with no older pupils. We therefore created two matched comparison groups: one for those participating schools that were included in the evaluation of Early Learning Goals, and one for those that were included in the evaluation of KS2 attainment. Most schools are included in the evaluation of both sets of outcomes.

We used 1:1 nearest neighbour matching based on propensity scores to create matched comparison groups. Schools were matched on the variables described in section 2.1.

The graphs in figure 2, known as love plots,[1] show how similar the participating and comparison schools were to one another, before and after matching, using a measure called the standardised mean difference. The mean difference is simply the difference between the average value of the variable for the participating schools, and the average value for the comparison schools. Standardising this measure means that we can compare balance across different variables. Generally, a standardised mean difference of 0.2 or below is considered to indicate good balance. This threshold is shown on the graphs as a dotted line.

As shown in figure 2, the matching process successfully created well-matched comparison groups. The +-0.2 boundaries are shown on the chart as dotted lines.

[1] Loveplots are named for Professor Thomas E. Love, who first developed them along with colleagues (https://academic.oup.com/eurheartj/article/27/12/1431/647407)

Figure 2: Standardised mean differences between participants and non-participants, before and after matching

4. Results

Results are given in several different forms: estimated impact, odds ratios, predicted probabilities, effect size, and months of progress.

In this report, we look at outcomes in four areas:

- Likelihood of achieving the Listening, Attention and Understanding (LAU) ELG
- Likelihood of achieving the Speaking ELG
- Attainment in KS2 reading
- Attainment in KS2 maths

The first two outcomes are binary; either a pupil achieves an Early Learning Goal (ELG) or they do not. We report the estimated effect on these outcomes using odds ratios. These ratios tell us the relative odds of a pupil achieving the relevant goal, depending on whether their school took part in the programme or not. An odds ratio of one would mean that a pupil in a participating school had exactly the same odds of entering as a pupil in a comparison school. An odds ratio above one means that a pupil in a participating school is more likely to achieve the ELG, and an odds ratio of below one means that they are less likely.

Odds ratios are not always easy to interpret. To aid with interpretation, we have also included the predicted probability of a pupil in participating school achieving the ELG and the predicted probability of a pupil in a matched comparison school doing so, for comparison.

The estimated impact on attainment in KS2 reading and maths are reported as raw scores. An estimated impact of one would suggest that we would expect a pupil in a participating school to achieve a score one mark higher than a pupil in a matched non-participating school.

We also include estimates of effect size for the KS2 outcomes. Effect size is a standardised version of the estimated impact. That is, it is the estimated impact divided by the standard deviation in the outcome measure. Because it is a standardised measure, it can be compared across different outcomes, so may be useful for comparing the magnitude of the programme's impact with that of other projects that have different outcomes.

However, effect sizes can be difficult to interpret; it is not immediately obvious whether an effect size of, for example, 0.5 is large or small. Months of progress are a measure used in education research to try and help with this. In this report, effect sizes were translated into equivalent months of progress using guidance developed by the Education Endowment Foundation, as shown in table 3.[2] In our example, an effect size of 0.5 would be the equivalent of six months of additional progress; expressed using the months of progress measure, it is clear that this is a large effect.

Table 5: Effect sizes and equivalent months of progress

Effect size from	To	Months of progress
-0.04	0.04	0

[2] https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluation/evaluation-guidance-and-resources/reporting-templates, *Evaluation report template*, accessed October 2023

0.05	0.09	1
0.10	0.18	2
0.19	0.26	3
0.27	0.35	4
0.36	0.44	5
0.45	0.52	6
0.53	0.61	7
0.62	0.69	8
0.70	0.78	9
0.79	0.87	10
0.88	0.95	11

4.1 Early Learning Goals

Overall

Estimates of the impact of participation in the programme on the likelihood of pupils achieving Early Learning Goals in listening, attention and understanding and in speaking are shown in the tables below, with 95% confidence intervals (all to two decimal places).

We report the estimated effect on this outcome using odds ratios. These ratios tell us the relative odds of a pupil achieving the relevant Early Learning Goal, depending on whether their school took part in the programme or not. An odds ratio of one would mean that a programme participant had exactly the same odds of achieving the goal as a comparison pupil. An odds ratio above one means that a participant is more likely to achieve the goal, and an odds ratio of below one means that they are less likely.

Table 6: Estimated effect of participation on likelihood of achieving Early Learning Goals

Outcome	Lower CI	Estimate	Upper CI	No. schools
LAU	0.94	1.05	1.19	810
Speaking	0.96	1.08	1.22	810

These results do not provide conclusive evidence that participation in the programme gas a positive impact on the likelihood of achieving these Early Learning Goals. While the point estimates for both outcomes are above one, indicating a positive estimate, the lower confidence intervals for both estimates are below one. This means that we cannot be confident that the programme has an impact on these outcomes.

The table below shows the predicted probabilities of pupils in participating schools and those in matched comparison schools achieving the relevant Early Learning Goals. These probabilities may be easier to interpret than odds ratios.

Table 7: Predicted probabilities of participants and pupils from matched comparison schools achieving Early Learning Goals

	Predicted probability		No. schools	
Outcome	Treated	Comparison	Treated	Comparison
LAU	79%	78%	405	405
Speaking	80%	78%	405	405

Years

Estimates of the impact of participation in the programme on the likelihood of pupils achieving Early Learning Goals in listening, attention and understanding and in speaking, broken down by length of participation, are shown in the tables below, with 95% confidence intervals (all to two decimal places).

Table 8: Estimated effect of participation on likelihood of achieving Early Learning Goals, by length of participation

Outcome	Years	Lower CI	Estimate	Upper CI	No. schools
LAU	0.5	0.85	1.07	1.38	168
	1	0.97	1.15	1.38	344
	1.5	0.66	0.86	1.16	144

	2	0.78	0.99	1.27	154
Speaking	0.5	0.87	1.11	1.47	168
	1	0.98	1.17	1.40	344
	1.5	0.69	0.90	1.24	144
	2	0.77	1.00	1.29	154

These results do not provide conclusive evidence of a positive impact for pupils who took part in the programme, although the majority of point estimates are positive. The estimated impact on those whose schools took part in the programme for one or one and a half years is slightly lower than that on those whose school had taken part for longer.

The table below shows the predicted probabilities of pupils in participating schools and those in matched comparison schools achieving the relevant Early Learning Goals. These probabilities may be easier to interpret than odds ratios.

Table 9: Predicted probabilities of participants and pupils from matched comparison schools achieving Early Learning Goals, by length of participation

		Predicted probability		No. schools	
Outcome	Years	Treated	Comparison	Treated	Comparison
LAU	0.5	80%	79%	84	84
	1	80%	78%	172	172
	1.5	78%	79%	72	72
	2	78%	78%	77	77
Speaking	0.5	80%	78%	84	84
	1	80%	78%	172	172
	1.5	79%	78%	72	72
	2	78%	78%	77	77

Dosage

Estimates of the impact of participation in the programme on the likelihood of pupils achieving Early Learning Goals in listening, attention and understanding and in speaking, broken down by dosage, are shown in the tables below, with 95% confidence intervals (all to two decimal places).

Dosage was determined based on responses to a feedback survey administered by Voice 21 to teachers in participating schools, specifically by a question on whether teachers were actively applying the guidance received from Voice 21 in their teaching. Schools in which less than 33% of teachers were doing so were defined as low dosage, between 33 and 66% as mid dosage, and 66% or over as high dosage. Unfortunately, relatively few participating schools responded to the feedback survey; those that did not are excluded from this part of the evaluation.

Table 10: Estimated effect of participation on likelihood of achieving Early Learning Goals, by level of engagement

Outcome	Dosage	Lower CI	Estimate	Upper CI	No. schools
LAU	Low	0.59	1.07	2.35	24
	Med	0.86	1.09	1.42	138
	High	0.77	1.01	1.40	98
Speaking	Low	0.60	1.11	2.47	24

	Med	0.86	1.10	1.46	138
	High	0.78	1.04	1.45	98

Again, these results do not provide conclusive evidence of an impact on pupils at any level of dosage, although the point estimates are all positive. There is no indication that pupils who attended a school with a higher level of engagement saw a stronger impact on these outcomes.

The table below shows the predicted probabilities of pupils in participating schools and those in matched comparison schools achieving the relevant Early Learning Goals. These probabilities may be easier to interpret than odds ratios.

Table 11: Predicted probabilities of participants and pupils from matched comparison schools achieving Early Learning Goals, by level of engagement

		Predicted probability		No. schools	
Outcome	Dosage	Treated	Comparison	Treated	Comparison
LAU	Low	83%	81%	12	12
	Med	80%	78%	69	69
	High	79%	78%	49	49
Speaking	Low	84%	81%	12	12
	Med	80%	78%	69	69
	High	79%	78%	49	49

Key Stage 2

Overall

Estimates of the impact of participation in the programme on Key Stage 2 attainment in reading and maths are shown in the tables below, with 95% confidence intervals (all to two decimal places). Also included in the tables are estimates of effect size and equivalent months of progress.

Table 12: Estimated effect of participation on attainment at KS2

Outcome	Lower CI	Estimate	Upper CI	Effect size	Months of progress	No. schools
Reading	-0.19	0.17	0.54	0.02	0	790
Maths	-0.17	0.18	0.56	0.02	0	790

These results do not provide conclusive evidence that participation had a positive impact on Key Stage 2 attainment in reading or maths. While the point estimates for both outcomes are positive, the confidence intervals for both include zero, meaning that we cannot be confident that the programme has an impact on these outcomes.

Years

Estimates of the impact of participation in the programme on Key Stage 2 attainment in reading and maths, broken down by length of participation, are shown in the tables below, with 95% confidence intervals (all to two decimal places). Also included in the tables are estimates of effect size and equivalent months of progress.

Table 13: Estimated effect of participation on attainment at KS2, by length of participation

Outcome	Years	Lower CI	Estimate	Upper CI	Effect size	Months of progress	No. schools
Reading	0.5	-0.69	0.04	0.86	0.00	0	166
	1	-0.35	0.23	0.83	0.02	0	324
	1.5	-0.55	0.25	1.11	0.02	0	152
	2	-0.73	0.02	0.86	0.00	0	148
Maths	0.5	-0.68	0.08	0.99	0.01	0	166
	1	-0.21	0.40	1.02	0.04	0	324
	1.5	-0.92	-0.11	0.78	-0.01	0	152
	2	-0.73	0.10	0.93	0.01	0	148

Again, these results do not provide conclusive evidence of an effect on any of the groups. However, almost all of the point estimates are positive. For the reading outcome, the point estimates are highest for those who had participated for one or one and a half years, and were lower for those who had taken part for a shorter or longer period. For the maths outcome, the point estimate was highest for those who had participated for one year, but similar for all other lengths.

Dosage

Estimates of the impact of participation in the prorgamme on Key Stage 2 attainment in reading and maths, broken down by dosage, are shown in the tables below, with 95%

confidence intervals (all to two decimal places). Also included in the tables are estimates of effect size and equivalent months of progress.

Dosage was determined based on responses to a feedback survey administered by Voice 21 to teachers in participating schools, specifically by a question on whether teachers were actively applying the guidance received from Voice 21 in their teaching. Schools in which less than 33% of teachers were doing so were defined as low dosage, between 33 and 66% as mid dosage, and 66% or over as high dosage. Unfortunately, relatively few participating schools responded to the feedback survey; those that did not are excluded from this part of the evaluation.

Table 14: Estimated effect of participation on attainment at KS2, by level of engagement

Outcome	Dosage	Lower CI	Estimate	Upper CI	Effect size	Months of progress	No. schools
Reading	Low	-1.44	0.09	1.79	0.01	0	28
	Med	-0.63	0.17	1.10	0.02	0	136
	High	-0.64	0.22	1.18	0.02	0	98
Maths	Low	-0.96	0.73	2.52	0.08	1	28
	Med	-0.61	0.24	1.22	0.03	0	136
	High	-0.41	0.54	1.47	0.06	1	98

These results do not provide conclusive evidence of an impact at any of the dosage levels, although, again, all of the point estimates are positive. However, the point estimates for the reading outcome are higher for schools with higher levels of engagement. For the maths outcome, the point estimate is highest for the low dosage group, but the confidence interval for this group is particularly wide, suggesting that this outcome is less reliable.

5. Conclusions

5.1 Overview

We did not find conclusive evidence that the programme had an impact on either Key Stage 2 attainment or on the likelihood of pupils achieving the relevant Early Learning Goals. However, while estimates of the impact were not statistically significant, they were positive for all of the outcomes measured.

We also looked at the impact broken down by length of participation in the programme, but did not find evidence to suggest that that the programme had a higher impact on those schools that had taken part over a longer period. Indeed, we found little consistent evidence of the impact varying by length of participation.

Finally, we looked the impact broken down by level of engagement with the programme. This analysis was more limited as data on level of engagement was unavailable for a number of participating schools. The results when looking at impact by level of engagement were mixed; we found some evidence to suggest that the programme has more of an impact on KS2 reading in schools that reported higher levels of engagement from their teachers, although this didn't appear to be the case for the other outcomes.

5.2 Limitations

This evaluation uses a quasi-experimental design; it relies on creating a matched comparison group that is statistically similar to the programme participants, based on data from the NPD. Creating a comparison group in this way means that we are unable to control for factors not recorded in the NPD, for example pupil motivation, social class or parental occupation.

Of particular importance for this evaluation, the NPD does not include data on prior attainment for pupils working towards Early Learning Goals, meaning that we were unable to control for prior attainment for the outcomes relating to Early Learning Goals.

Similarly, we are limited to outcomes to are recorded in the NPD. The programme aims to ensuring every child receives a high-quality oracy education, but the NPD does not include any direct measures of oracy skills. It may be the case that the programme has led to an improvement in oracy skills, but that this does not translate into an improvement in attainment at Key Stage 2, for example.

The programme has been running for a relatively short time and takes a whole school approach, largely working with school staff rather than directly with pupils. It may be the case that the programme will have an impact on the outcomes measured, but that this will not become apparent until schools have participated for a number of years. If this is the case, effects would not yet be measurable as the programme has only been running since September 2020.

The timeframe of the programme delivery includes the onset of the COVID-19 pandemic. This may have affected the delivery of the programme as well as the circumstances in individual participating schools, and mean that may the results of this evaluation not reflect the impact of the programme under pre-pandemic circumstances. However, this is a limitation common to most evaluations of programmes delivered during this period.