



What Works *for*
Children's
Social Care

Evaluation of Words for All

December 2022





What Works for Children's Social Care

Acknowledgments

We would like to thank all the schools who participated in the trial for supporting the evaluation during such difficult circumstances, as well as Professor David Waugh who provided invaluable advice on the process evaluation.

Funding

Funding for this evaluation report was provided by What Works for Children's Social Care. The authors declare no competing interests.

Authors

Doménica Àvila
Dr Mary Baginsky
Dr Andy Healey
Susannah Hume
Gabrielle McGannon
Dr Chiamaka Nwosu
Dr Carl Purcell

About What Works for Children's Social Care

What Works for Children's Social Care seeks better outcomes for children, young people and families by bringing the best available evidence to practitioners and other decision-makers across the children's social care sector. We generate, collate and make accessible the best evidence for practitioners, policymakers and practice leaders to improve children's social care and the outcomes it generates for children and families.

About King's College London

The Policy Institute at King's College London works to solve society's challenges with evidence and expertise. We combine the rigour of academia with the agility of a consultancy and the connectedness of a think tank. Our research draws on many disciplines and methods, making use of the skills, expertise and resources of not only the institute, but the university and its wider network too.

If you'd like this publication in an alternative format such as Braille, large print or audio, please contact us at: info@whatworks-csc.org.uk



Contents

Executive summary	4
Introduction	4
Objectives	4
Design	4
Findings	5
Conclusions	6
Introduction	7
Project background	7
Words for All	7
Methods	9
Impact evaluation	9
Implementation and process evaluation	10
Cost analysis	12
Findings	13
Impact evaluation	13
Implementation and process evaluation	19
Cost analysis	46
Limitations	50
Discussion	55
Conclusion	57
References	58
Appendix 1: Data tables	59



Executive summary

Introduction

In 2019, What Works for Children's Social Care conducted a re-analysis of data from 63 Education Endowment Foundation (EEF) randomised controlled trials (RCTs). In this re-analysis, the Vocabulary Enrichment Full Programme (VEFP), an intervention delivered by Bolton Local Authority in 2014, showed evidence of promise, specifically for pupils with experience of an intervention by local authority children's social care within the previous six years.

Since 2014, Whole Education, a network that brings together over 500 schools, trusts and education organisations, had further developed the VEFP into Words for All (WfA) in partnership with Bolton Local Authority. Therefore, in 2020, What Works for Children's Social Care commissioned an evaluation of Whole Education's updated intervention, WfA.

WfA offers training, materials and ongoing support to a "triad" of three teachers in participating secondary schools. Schools are then free to decide which elements of the intervention are best suited to their pupils and context.

King's College London evaluated WfA in the 2020/21 academic year. The evaluation included an impact evaluation, an implementation and process evaluation (IPE) and a cost analysis.

Objectives

Initially, four research questions were set for the impact evaluation. However, due to the small number of complete cases available for endline analysis, this was reduced to the following question:

1. What is the impact of WfA on the average reading skills (as measured by the New Group Reading Test) of Year 7 to 11 pupils who have been considered to be Children in Need, subject to a Child Protection Plan and/or Looked After in the past six years?

The implementation and process evaluation had the following aims:

- To evaluate the factors that facilitate and hinder the implementation of WfA, including the extent to which fidelity is maintained and, indeed, what constitutes fidelity
- To assess the confidence of teachers and other school-based professionals in the programme
- To understand the response of pupils to the programme.

The aim of the cost analysis was to estimate the cost per school of implementing WfA, as well as estimate a cost per pupil beneficiary.

Design

Impact evaluation

- The impact evaluation took the form of a two-sided randomised controlled trial, randomised at the family level and stratified by school



- Data lists were provided by 60 schools, with 2,748 pupils randomised at the beginning of the trial
- Eligible pupils were identified by schools and pseudonymised participant lists were provided to What Works for Children's Social Care for randomisation
- Originally, outcome data was to be collected by schools, with the New Group Reading Test (NGRT) and strengths and difficulties questionnaire (SDQ) to be completed by all eligible pupils at the beginning and end of the 2020/21 school year, under guidance from the research team
- Following low school compliance with testing requirements, some endline data was collected directly by the research team in October 2021.

Implementation and process evaluation

- The implementation and process evaluation took a mixed-methods approach with five stages of data collection:
 - November 2020: Observation of regional hub meetings
 - March–April 2021: Pre-intervention survey of participating schools
 - June–July 2021: Post-intervention survey of participating schools
 - June–July 2021: Interviews with staff from local authorities
 - November 2021: Case study interviews with participating schools.
- Due to COVID-19 restrictions it was not possible to proceed with initial plans to hold focus groups with pupils.

Cost analysis

- The cost analysis was based on intervention scenarios provided by Whole Education
- The analysis focused on estimating the opportunity cost of teacher time allocated to project-based work
- The cost analysis presents the cost of teacher time for specific project activities and the total cost for each project scenario on a per-school basis
- It also estimates the cost per pupil beneficiary for each project scenario by dividing the total cost by an estimate of the expected number of pupil beneficiaries per school.

Findings

- The RCT found no evidence that receiving WfA improved pupils' reading skills. The sign on the point estimate for the treatment group was positive, but it was not statistically significant ($p=0.33$). Substantial issues with data collection and quality mean that this finding is presented with low confidence
- The COVID-19 pandemic had a substantial impact on the ability of schools to engage with and deliver WfA. By March 2021, approximately one-third of participating schools had dropped out of the programme; this increased to almost half of schools by the end of the trial. Many of the schools that did implement the intervention reported significant difficulties due to COVID-19 restrictions and capacity problems
- Despite efforts to adapt the programme to the COVID-19 context – for example, pushing back the start date and reducing the intervention time to six weeks – many school and



local authority staff believed that the programme should have been delayed until the following year

- Due to the lack of structure in the WfA programme, the IPE found no evidence of fidelity to a programme, with schools taking significantly varied approaches to implementation. This included some schools that did not deliver the intervention for the minimum six-week period
- Many staff were satisfied with the support that had been provided by Whole Education, but some raised issues with engaging due to COVID-19 demands, while others felt the support had not been adequate, including some that had not had any contact from their Whole Education mentor despite efforts to engage
- Staff at both schools and local authorities were not sure that the cohort selected for the intervention (pupils with experience of social care in the past six years) was the most appropriate given the difficulties of restricting the programme to just some pupils in a year group or class, the range of academic abilities within the group and concerns about young people feeling targeted
- The cost analysis found that the programme cost per school ranged from £1,995 to £2,660 depending on the implementation scenario. The average cost per pupil beneficiary was between £81 and £94.

Conclusions

- This intervention and evaluation were delivered in very unusual circumstances, with limited fidelity to any kind of programme design, and significant limitations to the analysis. Therefore, it is not recommended that the evaluation findings are generalised to other populations or target groups
- As with the EEF RCT of VEFP, this trial did not find a significant effect of the programme on reading skills, despite detecting a small positive trend in the treatment group. Any future trials would require a very large sample to ensure that the trial is adequately powered to detect a potentially small effect size
- Low buy-in from schools was an ongoing issue for the evaluation. Though the impact of COVID-19 on schools' capacity should not be underestimated, it is recommended that greater stakeholder consultation takes place in future trials before the trial commencing. This may help mitigate some of the confusion in relation to the trial's design and rationale that contributed to poor data collection and quality in this trial.



Introduction

Project background

In 2019, What Works for Children's Social Care (WWCSC) conducted a re-analysis of 63 randomised controlled trials (RCT) funded by the Education Endowment Fund (EEF). Data from the trials was analysed to assess whether the interventions showed signs of promise when the cohort was restricted to young people who had experience of an intervention by local authority children's social care (CSC) within the six years before the beginning of the trial. The six-year definition was chosen to be consistent with the definition of free school meals eligibility status commonly used in education research. Ten interventions were identified in this exploratory process (Sanders et al., 2020), and three taken forward to trial with WWCSC's target cohort.

One of the interventions selected for trial was the Vocabulary Enrichment Full Programme (VEFP), a reading improvement programme that had been delivered to Year 7s in 12 schools across Bolton Local Authority in 2013/14. Though the EEF RCT did not find any evidence that the programme had an impact on reading attainment for the full cohort, the re-analysis suggested that it may have potential for pupils with experience of CSC.

To test the impact of the programme on CSC-experienced pupils, WWCSC commissioned Whole Education to deliver the Words for All (WfA) programme as part of an RCT focused on this cohort. WfA was designed to build on the learnings of the VEFP, as well as incorporating additional evidence on best practice that had been published since the original trial in 2013/14.

King's College London was commissioned by WWCSC to conduct a two-armed RCT to measure the impact of WfA on CSC-experience pupils' reading skills, and an implementation and process evaluation to understand the fidelity of programme implementation, as well as staff and pupil experiences of the programme.

Words for All

WfA was built on the VEFP, a combination of three existing programmes – the Vocabulary Enrichment Intervention Programme (VEIP), Sounds-Write and Literacy Plus – and aimed to improve the reading abilities of pupils in Year 7. It focused on literacy catch-up at the transition from primary to secondary school. In 2013/14, EEF commissioned the National Foundation for Educational Research to conduct an RCT alongside an implementation and process evaluation (IPE). The RCT involved 649 Year 7 pupils in 12 schools in Bolton Local Authority. The EEF evaluation found that VEFP did not have an impact on reading attainment (Styles et al., 2014).

Whole Education is a network that brings together over 500 schools, trusts and education organisations. Following the EEF evaluation of the Vocabulary Enrichment Full Programme, Whole Education worked with Bolton Local Authority to design a programme that built on the lessons learned from the 2013/14 trial. WfA is a collaborative programme, with a focus on



supporting secondary schools to design and implement vocabulary interventions that are targeted to each school's context.

For this trial, Whole Education worked with a group of three teachers (a “triad”) in each participating secondary school to select a range of materials and strategies that would work best in that school's context. The materials and strategies were to be delivered to pupils with experience of CSC in Key Stages 3 and 4 (Years 7 to 11).

To facilitate this, WfA offered the following support for teachers:

- Hub meetings bringing together triads from local schools
- A WfA coach to support triads through ongoing coaching calls
- Expert webinars
- School visits by Whole Education staff.

The intention was that the options would be taken as a guide for evidence-based practice, rather than prescriptive interventions.

Teachers could select from the following options:

- Reading Out Loud – treatment group pupils were withdrawn from class to participate in a session where the school librarian read out loud extended texts to pupils (e.g. a novel), rich with complex vocabulary, to expose pupils to new words and meanings
- Academic Vocabulary Building – treatment group pupils were withdrawn from class at the beginning of the week to be pre-taught pre and parallel instruction of vocabulary to support pupils to access new curriculum content, including exploring etymological roots to aid comprehension and application
- Academic Non-fiction – pre and parallel teaching of high-quality non-fiction texts to support curriculum units, to develop wider context and cultural understanding and to develop strategies for learning new vocabulary that pupils encounter
- Talking about Learning – support for pupils to articulate ideas and spoken expression through, for example, structured questioning to develop reading comprehension and to extend spoken vocabulary. Intentional and planned-for opportunities to talk
- Reading for Pleasure – explicit support and mentoring for pupils to find more enjoyment in reading by making informed recommendations (based on knowledge of age-appropriate literature) engaging in ‘book talk’ and creating space for independent reading time within a highly social reading environment.

The programme was launched in term 1 of the 2020/21 school year. At this time, it was anticipated that WfA would be a year-long programme, with the intervention delivered to pupils throughout the spring term. However, largely due to COVID-19, this timeline had to be adjusted, with the pupil-facing element of the intervention reduced to six weeks in the first half of the summer term (May–June 2021).



Methods

Due to the extent of agency each school was given in their implementation of WfA, the intervention measured by this evaluation was the establishment of triads in schools, and the support provided by Whole Education to develop and implement individual school programmes.

Impact evaluation

The impact evaluation was a two-armed RCT, randomised at the family level and stratified by school.

Research questions

The primary research question for the impact evaluation was:

1. What is the impact of WfA on the average reading skills (as measured by the New Group Reading Test) of Year 7 to 11 pupils who have been considered to be Children in Need, subject to a Child Protection Plan and/or Looked After in the past six years?

The secondary research questions were:

1. What is the impact of WfA on GCSE attainment (as measured by 1. Attainment-8, and 2. English Language GCSE grade) of Year 11 pupils who have been considered to be Children in Need, subject to a Child Protection Plan and/or Looked After in the past six years?
2. What is the impact of WfA on social and emotional skills (as measured by the strengths and difficulties questionnaire) of Year 7 to 11 pupils who have been considered to be Children in Need, subject to a Child Protection Plan and/or Looked After in the past six years?
3. What is the impact of WfA on number of sessions absent of Year 7 to 11 pupils who have been considered to be Children in Need, subject to a Child Protection Plan and/or Looked After in the past six years?

However, due to the reduced sample size available and issues with data quality (please see Limitations section of this report for more details) it was decided not to go ahead with secondary research questions 1 and 3. Further, due to a copyright issue, it is not possible to report findings from the strengths and difficulties questionnaire (SDQ).

Randomisation

Family-level randomisation was selected to avoid spillovers between siblings. All pupils who had been in contact with CSC in the last six years were eligible for the programme. Schools provided lists of eligible pupils (including a family identifier) and randomisation was conducted by WWCS, with balance checks on sex, year group and type of CSC experience. Schools were provided with treatment and control lists for programme delivery, and King's was provided with a pseudonymised allocation list of all randomised pupils.



Concerns were raised about the quality of the data lists provided by schools through the IPE, as well as the difficulties of avoiding spillovers between the treatment and control groups within each school. Further, issues with treatment allocation were identified during endline data collection. These concerns are addressed in the Limitations section of this report.

Sample size

At the beginning of the programme it was anticipated that 79 schools would participate. However, over the course of the programme almost half of the original 79 schools withdrew, largely due to COVID-19. Sixty schools provided pupil lists to WWCS, resulting in 2,748 pupils being randomised and, at the end of the 2021/22 academic year, Whole Education reported that 42 secondary schools had delivered the intervention, representing 2,157 of the randomised pupils. However, many schools were uncontactable for both the impact evaluation and the IPE, resulting in a substantially reduced final sample of 221 complete cases and 137 partial cases. This is discussed further in the Limitations section.

Data collection

The primary outcome was the Standard Age Score (SAS) obtained via the New Group Reading Test (NGRT). As part of WfA, participating schools were meant to complete the NGRT with all in-scope pupils at three timepoints (beginning of term 1, middle of term 2 and end of term 3). To inform the secondary research questions, schools were to deliver the SDQ alongside the NGRT at baseline and endline. As schools were already being funded directly by WfA to administer the NGRT, to reduce testing burden on pupils the SAS from the first and third timepoints were to be provided to King's for analysis.

However, as data was sought from participating schools it became clear that a significant number of pupils had not completed the NGRT at baseline (June/July 2020) or at endline (July 2021). For example, in one school where 57 pupils were randomised, data was provided for only 11. Therefore, in September 2021 it was agreed that King's would make efforts to collect data directly from schools where no endline data had been collected (for discussion of the differences between pre- and post-summer holiday data please see the Limitations section of this report).

The initial trial plan was that only schools who participated in both the baseline and endline IPE surveys would be included in the analysis for the RCT to allow for a dosage response indicator as part of the secondary analysis. However, due to low engagement from schools, it was decided to include all schools in the endline data collection to maximise available results.

Implementation and process evaluation

The IPE took a mixed-methods approach, including observations, surveys and case study interviews.

Research questions

The research questions for the IPE were designed to explore programme fidelity, and staff and pupil experiences of the programme.

1. What are the expectations placed on schools involved in the WfA programme?



2. What are schools' understanding of the essential components of the programme?
3. How have schools:
 - a. Adhered to the essential components of the programme?
 - b. Supplemented them?
4. What factors have hindered and/or facilitated the implementation of the programme?
5. To what extent do staff feel that their participation in the programme has improved their skills and confidence in identifying and responding appropriately to the needs of pupils?
6. To what extent do staff perceive the programme to have been effective in addressing the needs of pupils?
7. What are the experiences of pupils who have participated in the programme and what, if any, benefits did they perceive from it?

Given restrictions created by COVID-19, it was not possible to proceed with the planned methods to address research question 7. Therefore, the IPE findings presented in this report will only respond to questions 1 to 6.

Data collection

The original methodology was adapted to fit with COVID-19 restrictions. The implementation and process evaluation involved five stages of data collection:

- November 2020: observation of regional hub meetings organised by Whole Education to explain the requirements for participating schools and discuss potential strategies/schemes for individual schools to use as part of their intervention programmes. The researchers took detailed notes and Whole Education provided video-recordings of the hub events
- March–April 2021: pre-intervention survey of participating schools using the Online Surveys platform
- June–July 2021: post-intervention survey of participating schools using the Online Surveys platform
- June–July 2021: interviews with staff from local authorities. All interviews were recorded and transcribed in full
- November 2021: case study interviews with participating schools. All interviews were conducted virtually using Microsoft Teams or Zoom and were recorded and transcribed in full.

The original plan for the IPE included interviews with social workers and other professionals working with the children taking part in the programme. There were also plans to hold focus groups with the children. The pandemic restrictions ruled out the latter and contact with schools indicated that social workers and other professionals were unlikely to be aware of the programme. Instead, it was agreed that representatives of local authorities where the schools were based would be interviewed.

Charts presented from the IPE survey were generated by the Online Surveys platform. The qualitative data collected through the observations, open-ended survey questions and interviews were analysed using a reflexive thematic approach (Braun and Clarke, 2006).



Cost analysis

A cost analysis was undertaken using different project scenarios developed by the programme providers. These were intended to provide a broad characterisation of the types of projects that could be undertaken within schools. In each case it was assumed that each project would be undertaken over a six-week period. The scenarios do not reflect actual practice within schools participating in the randomised control trial: appropriate data on actual project activity and resource utilisation within participating sites was not collected as part of the wider evaluation and therefore actual project activity could not be costed. Instead, the provider identified scenarios describing three different types of interventions that might be undertaken: lesson-based work with eligible pupils; small group or one-to-one sessions carried out outside school hours (e.g. as part of existing extra-curricular tutorial time); and a home-based approach whereby parents lead on vocabulary work with children with initial briefing and training from triad teachers.

The provider was asked to identify the key activities required to support each type of initiative, including: support work from the programme provider (relevant to all project scenarios); interaction between lead staff members and other teacher triads (e.g. staff development and training); preparatory work before pupil interaction; and face-to-face interaction with eligible pupils. The provider then made an estimate of the quantity of teacher time (in hours) that would be required to support each activity. The cost analysis focused on estimating the opportunity cost of teacher time allocated to project-based work. 'Opportunity cost' is a key concept in economic analysis: in the context of this evaluation, it is intended to reflect the scarcity of teacher time invested in project-based work and the value of alternative uses of time forgone. As is standard practice in economic analysis, hourly pay (with additional costs associated with teacher employment) are used to approximate the value of an additional hourly 'unit' of teacher time invested in the WfA programme.

The cost analysis presents the cost of teacher time for specific project activities and the total cost for each project scenario on a per-school basis. It also estimates the cost per pupil beneficiary for each project scenario by dividing the total cost by an estimate of the expected number of pupil beneficiaries per school. For this estimate, the number of project beneficiaries was assumed to equate to the mean number of eligible pupils at each school randomised to the WfA programme in the RCT (with the minimum and maximum number of pupils at randomised schools used to provide a lower and upper bound estimate of the cost per beneficiary).



Findings

Impact evaluation

Research question 1: what is the impact of WfA on average reading skills?

Primary analysis

The analysis found no evidence that receiving WfA improved pupils' reading skills.

The primary research question for the RCT was the impact of WfA on average reading skills, as measured by the SAS produced by the NGRT. Complete data (NGRT results at baseline and endline) was available for 111 pupils assigned to receive WfA, and 110 pupils assigned to the control group.

Originally, an intention to treat (ITT) analysis was planned as the primary analysis for this outcome. However, due to concerns about non-adherence to the randomised assignment in some schools, the trial protocol was updated with a complier-averaged causal effect (CACE) as the primary analysis, with the ITT retained as secondary analysis. CACE analysis measures the results of participants who complied with their allocation to either the treatment or the control group, so it is a more appropriate measure where compliance with treatment allocation is low. In this trial, 17% of pupils assigned to the control group received the programme, while 24% of treatment-assigned pupils did not. This data was obtained by asking schools to provide lists of pupils who were included in the treatment and control groups after the intervention. These lists were cross-checked against the original randomisation to identify level of compliance. Table 1 presents the compliance rates across groups.

Table 1. Compliance rates

	Received control	Received treatment
Assigned to control	0.83	0.17
Assigned to treatment	0.24	0.76

On average, pupils who received WfA scored 5.14 points higher in the NGRT than control pupils, an effect that is not statistically significant (SD=5.29, p=0.33). Among the 221 available complete data cases (NGRT results at both baseline and endline), the average SAS across both baseline and endline examinations is 87.17. The CACE therefore represents a 5% improvement in reading skills, following the assumption that both groups would – in absence of the treatment – have followed the same trend. This effect is approximately one-fifth of a standard deviation in size – a small effect. However, due to the small sample size and compliance rates, this analysis lacks the statistical power to determine whether this effect is significant or causal. A full CACE analysis table can be found in Appendix 1. Figure 1 below shows the CACE treatment effect.

With a final sample of 221, and applying the assumptions used in the initial power calculations, the minimum detectable effect size was 0.3, adjusting for baseline outcomes. Hence, the effect



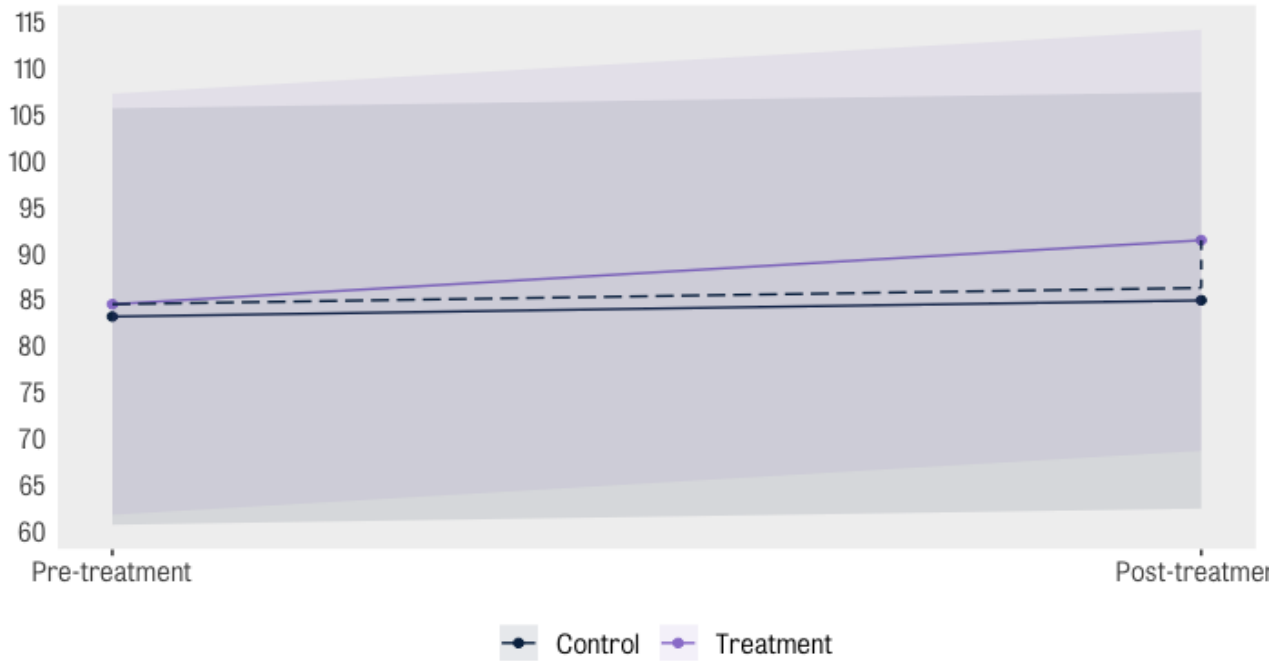
size of 0.218 standard deviations detected for the NGRT results falls out of the bound of detectability. Moreover, the issues with the compliance in this trial (please see the Limitations section for further details) mean that the statistical power is likely even lower. Therefore, a sample size of 221 is not large enough to robustly detect the estimated effect size under the conditions of this trial.

The power of an RCT refers to its capacity to capture a defined size effect. With a larger sample, RCTs can capture smaller changes in outcomes; when the sample reduces, the effect of the programme needs to be greater to be identifiable. While this RCT was initially powered to capture a small effect, during the implementation the design lost power for the following reasons:

1. The sample size was considerably reduced. This trial had significant levels of attrition associated with COVID-19 in both the implementation and the data collection. This meant that some schools did not deliver the programme, while others that delivered it either did not test the participating pupils or failed to share the data with the research team. There were several participants with missing baseline or endline data that prevented those observations from being included in the primary analysis
2. The programme effect became smaller. The spillovers of the programme into the control group translated into smaller differences between treatment and control groups. Also, as schools had autonomy over the delivery, the programme had different intensities across schools, which on average meant that lower-intensity schools pushed down the effects in higher-intensity schools. These issues shrank the detectable size effect, and a sample size of 221 could not compensate for this reduction
3. There was imperfect compliance with the treatment. Power calculations are done on the basis of perfect compliance, so the initial minimum detectable effect size of 0.3 for a sample of 221 assumes perfect compliance. As presented in Table 1, only 76% of those assigned to treatment took up WfA. When there is low take-up of the treatment, the sample size needs to increase to capture the same minimum detectable effect. As it did not, the RCT lost power.



Figure 1. CACE treatment effect



Pupils who received WfA (in purple) performed higher in the NGRT than their peers in the control groups. However, this difference is not large enough to push treated pupils' results out of the grey zone which represents the standard error. This graph therefore illustrates either that the programme has not changed pupils' reading skills, or that any effect is not identifiable with the current sample size.

CACE relies on three assumptions to capture the programme effect: stable unit treatment value (SUTVA), exclusivity criteria and monotonicity. The evidence suggests that these three assumptions may not have been met during programme delivery.

SUTVA assumes there were no spillover effects between the treated and control pupils. In this case, almost one-fifth of the staff members interviewed as part of the IPE acknowledged spillover effects across pupils. As randomisation occurred at the family level, and the programme involved upskilling teachers in evidence-based teaching methods, it is plausible that the control group pupils received some teaching influenced by the programme. Further, not all schools participated in the IPE, so it is possible that schools that did not participate in the IPE surveys and interviews had different experiences in relation to managing spillovers (please see the Limitations section for further discussion).

The second and third assumptions relate to the compliance conditions. CACE reflects the programme effect when it captures the pupils whose outcome changed only because they received the programme; these pupils are labelled the *compliers*. However, in every trial, the sample also includes other types of behaviours: always takers, defiers and never takers. The



latter two, *defiers* and *never takers*, could bias the effect estimation. Defiers are those participants who would always do the opposite of their treatment allocation. Similarly, never takers are participants who would never have engaged with WfA, even if assigned to treatment. The second and third assumptions therefore rely on providing evidence that CACE captures only the behaviour of the compliers.

Table 2 summarises the proportions of compliers, potential defiers and potential never takers in this trial.

Table 2. Compliers, potential defiers and potential never takers

	Proportion of the sample
Compliers	79%
Non-compliers	21%
Potential defiers	21%
Potential never takers	53%
Potential always takers	0.47%

The second assumption, exclusivity criteria, requires that the treatment effect is zero for those who didn't receive the treatment. This group covers 118 pupils: those who complied with the control assignment, and the non-compliers to the treatment assignment. In this case, it is not sensible to test the exclusivity criteria for two reasons: first, because the sample is too small to rigorously repeat the primary analysis on 118 pupils; and second, because the CACE analysis didn't find a significant positive effect in the first place. Therefore, it is not possible to ensure that the second assumption holds.

The third assumption, monotonicity, requires that there are no defiers – i.e. participants whose behaviour always opposes their treatment assignment, regardless of the assignment. As this assumption is not formally testable, and given the difficulties in the implementation of the trial due to COVID-19, it is not possible to gauge the extent to which the 41% of non-compliers (both in the treatment and control) may or may not be defiers.

Secondary analysis

The intention to treat (ITT) effect of the programme was also analysed. The ITT also found no statistically significant differences in reading skills between pupils assigned to the treatment and control.

ITT analysis is based on a participant's treatment assignment, regardless of whether participants comply with the assignment. On average, pupils assigned to WfA scored three points higher in the NGRT than control pupils. This is a small effect of 12.7% of a standard deviation. This effect is not statistically significant ($p=0.32$), including when controlled for variation by sex and year group, school fixed effects, clustered errors at family level or initial differences across pupils.



Given sample size constraints, this result suggests two possible scenarios: first, that assignment to receive WfA or not had no impact on pupils' reading skills; or, second, that this RCT doesn't have enough power to capture the effect of being assigned to receive the programme. A comparison CACE and ITT table is available in Appendix 1.

Sensitivity checks

As the final sample size meant that the trial was under-powered, several sensitivity checks were conducted to ensure that the results were not driven by noise. The CACE and ITT analysis was re-run, controlling for school fixed effects and/or year and sex covariates.

The findings of the sensitivity checks support the absence of a statistically significant effect. Both the CACE and ITT consistently maintained their sizes and positive direction across all the checks. Tables for the sensitivity checks for both the CACE and ITT are available in Appendix 1.

Given the small sample size, the programme's heterogeneous effects by sex or year group were not measured because this would require further slicing the data.

Imputing missing baseline

To maximise the available data, the primary analysis was extended to incorporate pupils with only endline records. A single-value imputation was used to generate baseline data. Two imputation methods were used. First, all missing baseline scores were imputed with the average endline SAS score (including both complete cases and cases with only endline data). Second, missing baseline results were replaced with the corresponding pupil's endline score. The two models of imputed results were compared to the original CACE and ITT NGRT results to identify changes in the coefficients.

The comparisons showed that the coefficients remained not significant, and there were changes in the detected effect sizes. When imputing missing baseline with the SAS mean, both the effect coefficient and the standard error shrank. This is true for the two imputed models. This analysis supports the hypothesis that attrition and low compliance with the treatment assignment prevented statistically significant results.

Full comparison tables for the CACE and ITT are available in Appendix 1.

Exploratory analysis: dosage indicator

To test whether higher exposure to WfA was associated with a larger improvement in reading skills a dosage indicator analysis was conducted. The indicator was developed from the information provided by staff members in the IPE. Staff members were asked the following question:

“On average, how many additional hours of literacy support do you estimate that each individual pupil received during this six-week period over and above what they received previously?”



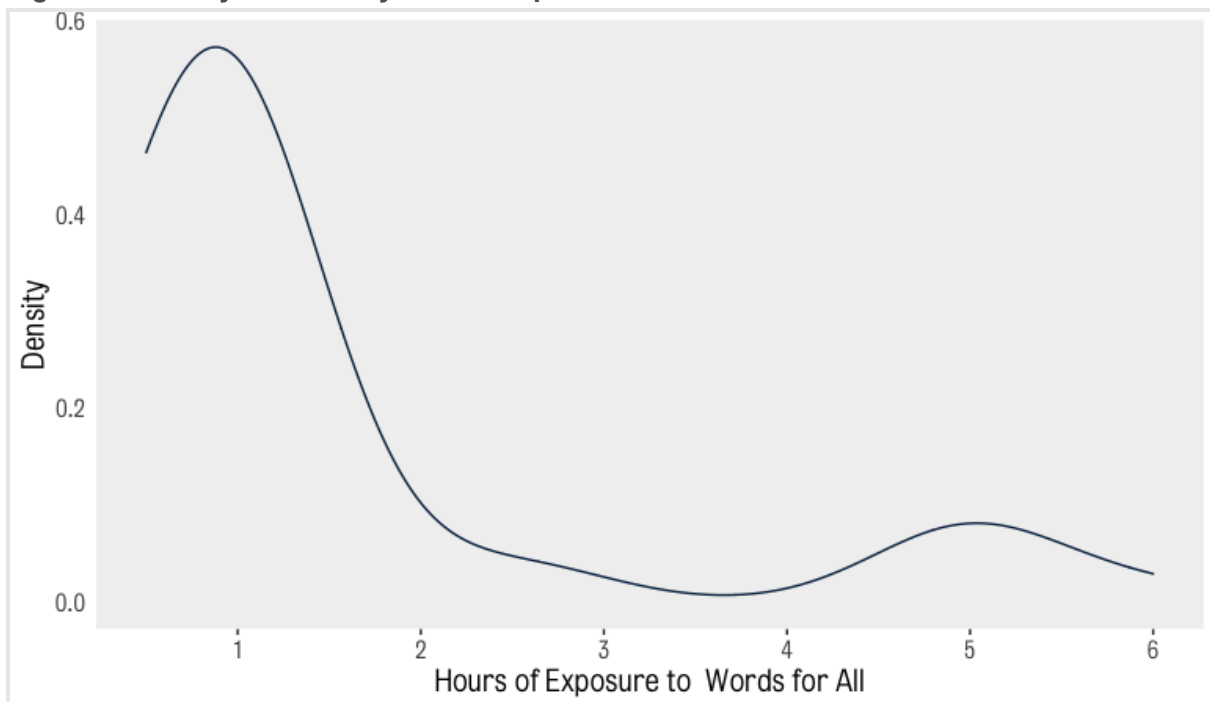
It should be noted that this was an open response question, and therefore not all responses were able to be used for the dosage indicator because some did not indicate a number of hours. Further, as noted in the IPE section, it is possible that some respondents misunderstood this question and provided answers based on the number of additional hours per week, rather than over the six-week period. These findings should therefore be treated with caution. The hours of additional support were combined with the baseline SAS results of 97 pupils from six schools that had responded to the survey. On average, pupils received 1.5 extra hours of literacy support per week.

To preserve the sample size, the dosage indicator analysis was based on pupils' assignment to treatment and control rather than their compliance. An exploratory model was constructed to test the association between the NGRT results and the dosage indicator. The model controlled for sex and year group, and clustered standard errors at family level.

Consistent with the findings in the sections above, it was not possible to detect a significant effect of the programme, regardless of the intensity of the treatment. The model did not find a statistically significant association between higher levels of exposure to WfA and improved results on the NGRT. A full results table can be found in Appendix 1.

To ensure that the results were not influenced by extreme or over-represented values, the distribution of the indicator was visualised, as shown in Figure 4.

Figure 2. Density of intensity of WfA exposure





As the distribution is not normal, the variable was transformed into a dummy with high intensity and low intensity levels. Using the mean value as reference, 80 pupils were classified as in the low intensity category, and 17 in the high intensity. It is important to note that the low intensity level includes all pupils in the control group. Consistent with the other findings, the transformed dosage indicator also did not return a significant result. A full table of results can be found in Appendix 1.

Implementation and process evaluation

Observation of regional hubs (November 2020)

Initially, 79 schools were recruited to participate in the trial by Whole Education at the beginning of the 2020/21 academic year. The original plan was for Whole Education to then work with school “triads” during the second half of the autumn 2020 term to develop bespoke school project plans and ensure that the requirements for participating in the trial were fully understood. Towards this end Whole Education organised a series of ten online regional seminars during November 2020. The researchers observed five of these seminars.

Across the 5 seminars observed attendance was low, with only 18 of the 38 invited schools present. The highest attendance was six of ten invited schools and the lowest was one from seven. Furthermore, for most schools only one member of staff was in attendance. It was clear from comments made by those who did attend that many schools were struggling with staffing capacity owing to absences and COVID-19 pressures, and that the timing of the seminars (during the school day) was also not ideal. It was apparent that in some schools their headteachers or other senior leaders had agreed with Whole Education to participate but then delegated responsibility to the staff in attendance. This was reflected in some instances by limited prior knowledge of the objectives and requirements of participation in the trial.

During the seminars Whole Education went over the timeline and requirements for the trial.

At this stage the plan was:

- Whole Education would work with schools individually and through the regional hubs during the second half of the autumn term to agree project plans by the end of the autumn term (18 December)
- Schools would consider basing their intervention programme on one of the five programmes or a combination of them. The five programmes recommended by Whole Education were Reading Out Loud, Pre-teach Vocab, Pre/parallel teaching of high-quality non-fiction text, Talk about Learning and Reading for Pleasure
- Schools would submit a ‘data list’ of eligible pupils (those who have, or have had, a social worker over the past six years) so that WWCS could allocate pupils to intervention and control groups before the end of the autumn term
- Schools would collect baseline data on their cohort of pupils by the end of the autumn term (18 December 2020) using the NGRT and SDQ



- (At that point) The intervention was to be carried out during the spring term 2021 (January–March), with hub meeting continuing and a series of “expert seminars” to be offered to triads to attend
- (At that point) Final endline data for NGRT and SDQs to be collected during the summer term (May–July 2021).

It is important to emphasise that participating schools were afforded significant autonomy in terms of the content of their intervention programmes. Schools were free to follow or adapt existing evidence-based interventions (including but not limited to the five recommended by Whole Education) or experiment with their own approaches. Moreover, it was evident from the group discussions, and subsequently from an examination of the project plans that were

submitted, that, although project plans were in the very early stages of development for most schools, intervention programmes across schools would vary considerably.

Pre-intervention survey (March–April 2021)

As noted earlier, the original plan had been for schools to have their project plans in place by the end of December 2020 so that they could implement their intervention programmes during the spring term (January–March). But, in the context of COVID-19, it became clear that this timeline was unrealistic and that schools needed more time to prepare their plans. Furthermore, in January 2021 all schools were closed to most pupils as part of government measures to tackle COVID-19. It was therefore decided that schools should continue their planning into the new year and that intervention programmes could begin in March/April 2021 (subject to pupils being allowed to return). It was later decided that intervention programmes must start by the week beginning 3 May 2021 and last for a minimum of six weeks (the duration of the first half of the summer term).

To coincide with the initiation of intervention programmes schools were asked to complete the pre-intervention surveys between mid-March and the end of April 2021. The purpose of the survey was to capture data relating to school project plans, staff resources, expectations for the intervention programme and overall experience of WfA ahead of the initiation of intervention programmes.

The pre-intervention survey was distributed to 55 schools; approximately one-third of schools had dropped out by March 2021 owing to COVID-19 pressures. More schools subsequently dropped out and only 30 completed the survey by the end of April 2021.

“Data lists” of eligible pupils

An essential requirement of participation in the trial was the submission of a data list of eligible pupils (those who had had a social worker in the past six years) to WWCS. Schools were subsequently notified which pupils had been randomly allocated to the intervention and control groups. Most – 28 of the 30 schools – had submitted their data list by the time they completed the survey.



Most survey respondents did not report difficulties in compiling their data list. However, a note of caution is needed as survey respondents indicated that they relied on colleagues in school (mainly safeguarding or inclusion leads) to provide the data and may not therefore have any great awareness of how it was compiled. Among the small number of respondents who did report difficulties, two commented that they had sought help from local authority colleagues whom they had not found to be responsive. One respondent questioned the number of pupils on the data list commenting: “a school of our size should have more pupils in this key group”.

Baseline New Group Reading Tests (NGRTs)

All schools were asked to complete a baseline NGRT with pupils in both the treatment and control groups. However, of the 30 schools that completed the pre-intervention survey, only eight schools said this had been done, 21 schools indicated that it was still to happen and one did not respond. Difficulties and concerns regarding the use of NGRT were raised, including difficulties in engaging older pupils, timetabling, IT problems and one school reported that it did not receive login details in time. Respondents from four schools highlighted specific difficulties in administering the test during COVID-19, including maintaining “bubbles” and an insufficient number of laptops.

Staff resources

The survey asked schools to provide background details for the three members of the triad who would lead the intervention programme and about any plans to involve other members of staff.

Key findings were:

- The majority of “first” triad members were English teachers or had a specific designated responsibility for literacy. A minority were special educational needs coordinators (SENCOs) or part of a learning support team. Twenty-nine had a teaching qualification (one did not respond)
- Among the 30 second triad members 16 were qualified teachers. The others included teaching assistants, pastoral workers and librarians
- Among the 25 third triad members 15 were qualified teachers (5 schools did not respond). The others included teaching assistants, special educational needs (SEN) support staff, librarians and one vice principal.

In two-thirds of all schools the plan was for other members of staff to be involved in implementing WfA. Where respondents were able to specify, it ranged from 1 to 12 additional members of staff. Most commonly the plan was to involve only English departments, but some schools planned to involve other departments including humanities (three schools), SEND (four schools), maths (one school) and science (three schools). Two schools said they were asking all staff/departments to be involved.

Support for the development of project plans

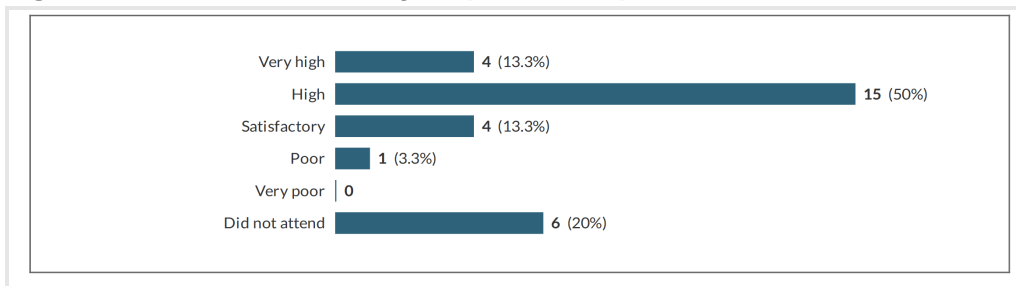
Ahead of the initiation of intervention programmes schools were required to work with Whole Education to develop and agree a project plan. As noted earlier, under the original timeline for



the project these should have been submitted by the end of December 2020. This was subsequently postponed to early May 2021 in line with the revised timeline.

For many staff their attendance at an online launch event was their first introduction to the WfA trial. The survey asked respondents to rate the usefulness of this event; 19 of the 30 respondents rated the event either highly or very highly, 4 as satisfactory and 1 rated it as poor. Six of the survey respondents had not attended the event.

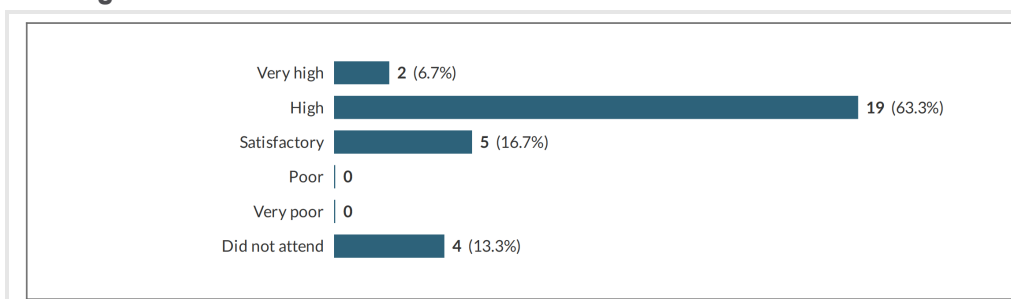
Figure 3. Baseline staff survey responses to question on usefulness of launch event



In line with these ratings additional comments were generally positive. Comments included: “engaging, really helpful strategies” and “excellent working knowledge of the project delivered”. Those who did not attend said that this was because they had not had the time. However, one respondent felt more negatively about the launch, commenting that “the aims of the project were not clear from the start”.

The survey also asked schools to rate the usefulness of the regional hub meetings in terms of helping to develop ideas for the project plans. Twenty-one of the 30 respondents rated them as either high or very high in this regard. Five respondents rated them as satisfactory and four had not attended the hub meetings.

Figure 4. Baseline staff survey responses to question on usefulness of regional hub meetings

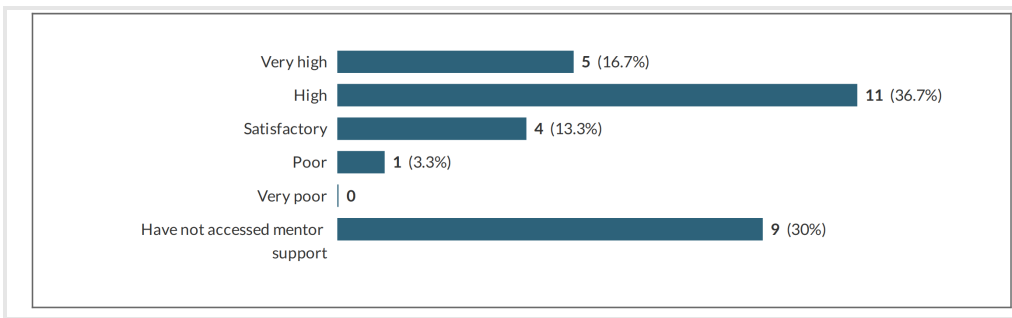


Additional comments highlighted difficulties some schools faced in terms of finding the time to participate in the meetings. It was noted earlier that attendance at the hub meetings observed by the researchers was low.



Schools were also asked to rate the usefulness of support provided by their school mentor allocated by Whole Education. Sixteen of the 20 respondents rated this as either high or very high. Four rated mentor support as satisfactory and one as poor. However, nine respondents had not accessed any support from their mentor.

Figure 5. Baseline staff survey responses to question on usefulness of support provided by school mentor



Among the schools that had not accessed mentor support three acknowledged in their additional comments that they had not had the time to link up. However, three schools commented that they had received no response from their mentors after they had contacted them.

Schools were also asked to provide more general comments regarding their experience of developing their project plans:

- Five schools highlighted difficulties in finding sufficient time, with three emphasising COVID-19-related pressures
- One respondent commented on difficulties engaging with colleagues in their own school
- One school highlighted the challenge of developing a programme for this cohort given the range of abilities and year groups.

This comment below from one school captured the challenges of developing project plans in the context of COVID-19:

“I think 2020/2021 has been an unfortunate time to run the project as there are so many immediate demands on people’s time and absence in staff and students is and has been an inevitable problem hampering the project. It has been useful to discuss and develop ideas with the team but when we first signed up I had hoped it had been more about being given a strategy to implement and analyse the effectiveness of it rather than coming up with our own projects.”

Further comments

Twelve schools provided further comments about their experience of the WfA trial to date. Two of these conveyed positive perceptions and an enthusiasm to get started with school interventions:

- *“We are really excited to see how this project can have an impact on our students”*



- *“It feels rewarding to take part in a project that has pupils’ outcomes at the forefront of its aims. That’s why we all became educators.”*

However, the remaining comments highlighted difficulties and concerns regarding the programme. Six of these schools highlighted a lack of capacity within school, particularly in the context of COVID-19, as the main obstacle they had faced. But four schools felt that they had not received enough support to develop their programmes:

- *“It has been very difficult to navigate due to lack of support from many sides”*
- *“It felt difficult following the hub meeting to gain full insight into what and how we were to proceed. I didn’t always feel that the support was there when I needed it. I was a little disappointed by this”*
- *“I was thrust into the programme part way through so have found the process fairly stressful initially. I wonder about the variables in a group of people all coming up with their own bespoke plans. In my view it is hard for us to really pin-point if it is our project that has made the difference as you cannot compare it to how it has worked across a series of schools, hence the reason I would have preferred to be given a strategy to implement rather than come up with one for my school that differed from what everyone else was doing”*
- *“Would have preferred to have all the training sessions before putting together our plan.”¹*

Post-intervention survey (June–July 2021)

The post-intervention survey was distributed in the last week of June 2021 to the 30 schools that had completed the pre-intervention survey, by which time schools should have completed a period of implementation lasting a minimum of six weeks. The survey remained open until the end of the summer term.

Of the 30 contacted, 25 schools returned the survey. One school responded to say that it had dropped out of the trial for “various reasons”. Four schools did not respond to repeated requests to complete the survey and could therefore no longer be included in the trial.

Project plans: length of intervention

Of the 25 schools that completed the post-intervention survey 22 provided start and end dates for their programme. The table below shows the variation in the length of programmes across the schools. For schools that did not respond, and for those providing an autumn term start date, dates provided in the pre-intervention survey were used to calculate the length of programme. School holidays were also subtracted.

Table 3. Length of intervention delivered in schools that responded to post-intervention survey

Project length	Number of schools
----------------	-------------------

¹ Additional training was offered to schools in the form of “expert webinars”, but these were scheduled during the period when schools were implementing their intervention programmes.

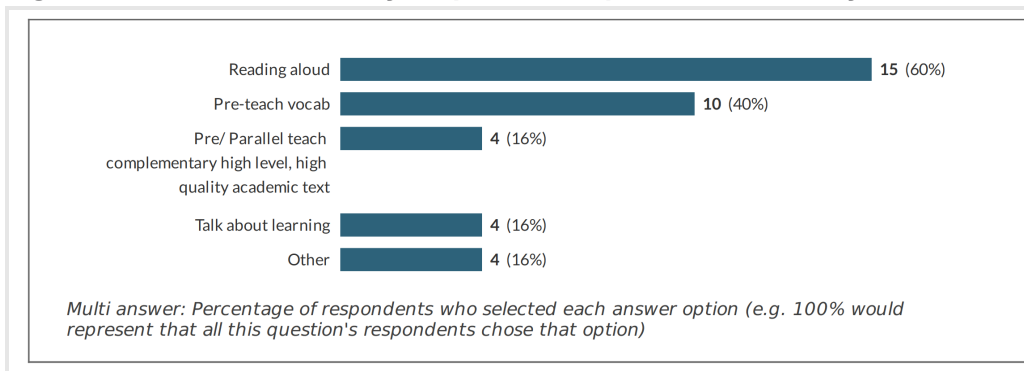


Less than 6 weeks ²	3
6–10 weeks	13
11–15 weeks	6
16+ weeks	3

Project plans: literacy interventions used

Schools were asked which, if any, of the four literacy interventions recommended by Whole Education³ their WfA programme would be based upon. More than one such programme could be used, as well as any other programme. Fifteen schools had used or adapted Reading Out Loud; ten had used Pre-teach Vocab; four had used Pre/parallel teach; and four had used Talk about Learning.

Figure 6. Endline staff survey responses to question on literacy interventions used



The four “other” programmes used were:

- Lexia
- SRA reading scheme and dialogic talk scheme
- Phonics
- Bedrock vocabulary academic reading.

Project plans: additional hours of literacy support

Schools were also asked how many *additional hours* of literacy support pupils on the programme received in total *over the first six weeks* of the programme (or the entirety of the programme if it only lasted six weeks). Twenty-three schools responded but only 20 provided a figure.

Table 4. Endline staff survey responses to question on additional hours of literacy support provided

Additional hours over 6 weeks	Number of schools
Less than 3 hrs	9

² This was less than the minimum period required.

³ Reading for Pleasure was dropped from the original list of five recommended interventions.



3–6 hrs	9
Over 6 hrs	2

It is possible that some schools may have misunderstood this question, perhaps providing a weekly figure.

Among the three schools that did not provide the number of hours support provided:

- Two schools explained that their programme could not be delivered as planned owing to pupil and staff COVID-19 isolations
- One school stated that “very little” additional support was provided “due to the low attendance and high exclusion rates seen in the intervention group”.

NGRT baseline and endline testing

All schools were asked to complete the NGRT at baseline and again after the intervention had been delivered. As noted earlier, at the time of the pre-intervention survey only eight schools confirmed that the baseline NGRTs had been completed. The post-intervention survey therefore asked schools about both the baseline and outcome surveys. The key findings were:

- Of the 25 schools, 22 had completed baseline assessments for pupils in the intervention group and 21 had completed them for the control group
- One of the three schools that had not completed the baseline assessments said that it had used its own tests to assess reading age
- Only ten schools had completed the endline assessments for pupils in their intervention group
- Only seven schools had completed the endline assessments for pupils in their control group
- Most schools who had yet to complete the endline assessments were planning to do so before the end of term or early in the autumn term.

Contact with schools in the summer term found that at least 11 schools had not completed the endline NGRT at all in the 2020/21 school year. Due to the difficulty in contacting schools to collect this information, it is possible that this number was higher.

In the comments almost half of schools highlighted challenges in administering the NGRT tests. Three commented on technical difficulties in setting up the test. However, concerns primarily related to the difficulties of timetabling the tests, which had been made more difficult owing to COVID-19-related challenges.

Asked about the appropriateness of the NGRTs for this project, six schools commented on the value of the data they produced. However, two schools felt that it was not appropriate for such a short intervention. Two further schools commented that it was not appropriate for all pupils to be asked to complete an NGRT, with one describing how it had left one Year 9 pupil feeling very distressed owing to the difficulty of the first few questions.



Schools' response to questions relating to the NGRTs also revealed some discrepancies in the sizes of the intervention and control groups between what had been reported in their responses to the survey and the allocations made by WWCS. These discrepancies were discussed with the schools who participated in the case study interviews, the findings of which are reported later in this report.

SDQ baseline and endline testing

Schools were also asked about the completion of baseline and endline SDQs for pupils in both the intervention and control groups, another requirement of the trial. The key findings were:

- Only 17 schools had completed a baseline assessment for pupils in their intervention group and only 11 had completed it for their control group
- Only five schools had completed endline assessments for pupils in their intervention group and only four had done so for their control group.

In relation to the administration of the SDQ:

- As with the NGRTs, several schools highlighted administrative difficulties relating to COVID-19
- Two schools commented that links to the SDQ were received late and after the NGRTs had been completed, thereby creating further administrative challenges
- One school forgot to ask pupils to complete it
- One said they did not realise this had to be done at all
- One school did not know this needed to be completed by pupils in the control group.

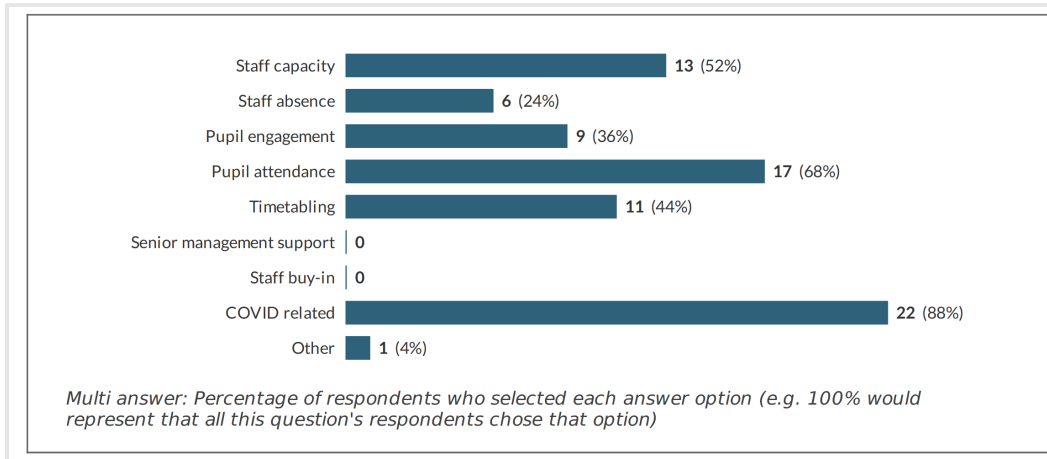
Asked about the appropriateness of the SDQ most schools did not raise any concerns. However, three schools felt that it was either not appropriate or not necessary for this intervention, with one commenting that some of the questions had made pupils feel uncomfortable.

Implementation difficulties

Schools were asked to select from a list of the main implementation difficulties they had faced.



Figure 7. Endline staff survey responses to question on implementation difficulties experienced



As the above chart shows, 22 of the 25 schools highlighted COVID-19-related issues, and over half highlighted staff capacity and pupil attendance. The school selecting “other” highlighted “exclusion related issues” and “lack of mentoring”.

When asked what they might have done differently with the benefit of hindsight 13 schools provided relevant comments. Eight schools said they would like to have had more capacity to develop and implement the programme and that COVID-19 challenges were a significant constraint. Four schools commented on internal changes they would make:

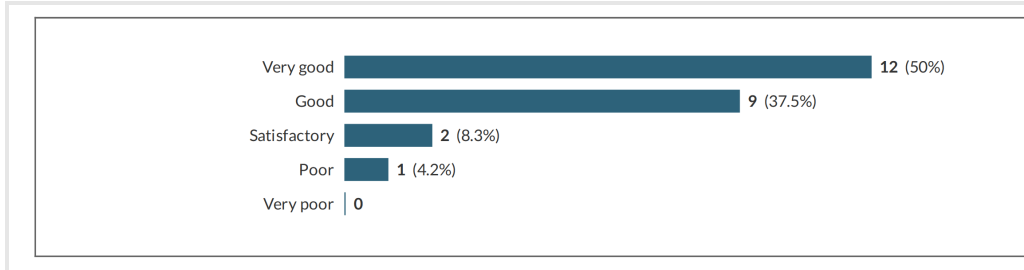
- *“I would have run the intervention in line with my department rather than inclusion as the different time of day for each year group made supporting staff and liaising as well as being present myself during the sessions very difficult”*
- *“Start the programme earlier, have the confidence to get straight on with it”*
- *“More time planning – whole project out with clearer control group”*
- *“I would involve more colleagues in the delivery of sessions so as to differentiate between pupils’ levels of ability. I would also build stronger links with carers/guardians so as to run the programme remotely for pupils who could access the internet from home. I would also incentivise pupils’ performance through a structured rewards system.”*

Support for schools

Schools were asked to rate the accessibility and effectiveness of support provided by Whole Education for the development and implementation of their WfA programmes. Twelve schools rated this as very good, nine as good, two as satisfactory and one rated it as poor.



Figure 8. Endline staff survey responses to question on support received from Whole Education



Additional comments were generally very positive about the support provided by Whole Education staff and through the expert seminars. For example:

- *“I think the webinars have been excellent”*
- *“Very supportive expert team”*
- *“Networking and CPD both excellent”*
- *“The session with X, Y and other participants were helpful prior to the project starting to help me formulate my ideas.”*

Two schools acknowledged that dealing with COVID-19-related challenges may have prevented them from getting more out of the programme. One commented:

“Some excellent resources and the people are brilliant. I just wish I could have accessed more of it and shared the information. Unfortunately, with other extreme demands placed on us it was just too much. Such a brilliant idea, I am sorry we were not able to take full advantage.”

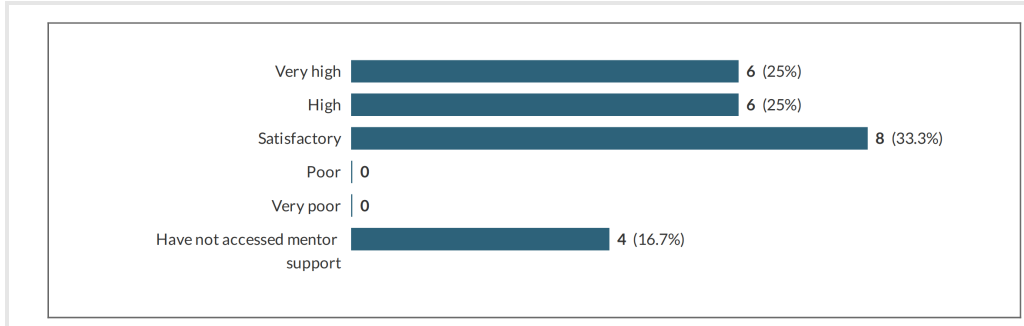
However, one school explaining their “poor” experience of the programme commented:

“I did not receive any communication from my mentor until during the Easter break to request my plan. I then received no response. The only support I have received has come from X who has always been brilliant and very helpful. The CPD sessions were great but unfortunately with TAGs and three exam groups it has not been possible to attend them live. There has been a significant delay in the sharing of any recordings after these sessions, making it more challenging and unmanageable to catch up on the content after the event.”

However, when asked to rate the usefulness of support provided by their school mentor to develop and implement their project plans schools provided mixed responses. Twelve schools rated this as high or very high, but eight rated it as only satisfactory and four had not received any support from their mentor.



Figure 9. Endline staff survey responses to question on usefulness of support received from school mentor



Eleven schools provided additional comments. Four spoke very positively, commenting:

- *“X was very helpful and supported me and any queries I had. Y was also extremely helpful and replied to any of my queries quickly”*
- *“The mentor was warm and approachable, and validated the project through her approval. She proposed ideas about data collection in the early stage of planning”*
- *“Excellent help and support”*
- *“Really enjoyed working with her.”*

An additional four schools acknowledged that they had not been able to utilise their mentors owing to competing demands on their time. However, three schools were more critical, commenting:

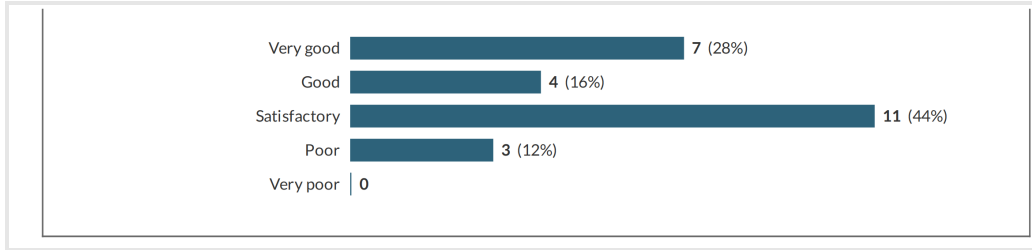
- *“No contact, bar requesting my plan, which I sent through and then heard no more”*
- *“The support given was fine but minimal. The Zoom session was useful and we had lots of insight. But there wasn’t a huge amount of contact”*
- *“We did not have mentor support until well into the intervention programme.”*

Peer support and learning

Mixed responses were also provided when schools were asked about the extent to which WfA had helped to develop and sustain a culture of peer learning and support within and between schools. Eleven schools rated this aspect of the programme as either good or very good, while 11 rated it as satisfactory and three as poor.



Figure 10. Endline staff survey responses to question on culture of peer learning and support



Twelve schools provided further comments. Three schools spoke positively about this aspect of the programme, commenting:

- *“I was able to liaise with three other schools about Words for All and now have contacts with other schools”*
- *“X should be highly commended for her thorough knowledge of practitioners and good practice, which she readily shared. Whole Education helped to forge links between peers through hubs and discussion following expert CPD, which permitted further refinement to the project and whole school strategies”*
- *“X was always available with support, encouragement and advice.”*

Five schools highlighted capacity problems and the demands of dealing with COVID-19 as a constraint on this aspect of the project. Three schools provided more negative comments.

- *“I have been put in direct contact with X which has been really worthwhile ... This was the only peer support we accessed”*
- *“Heard comments and plans on Zoom calls, but interaction has not gone deeper than that. Perhaps ‘buddying up’ schools would work?”*
- *“Apart from the remote meetings, very little contact with other schools.”*

Cohort

Schools were also asked whether they thought the cohort for this intervention – pupils who have had a social worker in the past six years – was the right one and the responses fell into two equal groups: 12 schools said yes and 13 said no.

Two key concerns were highlighted by the 13 schools asked to explain why they did not think this was the right cohort. Four schools highlighted how difficult it had been to engage pupils facing a range of difficulties including some who are poor school attenders.

For example:

- *“From feedback, most of these students do not want to be targeted due to their individual circumstances that are not within their control”*
- *“Many of the students selected were non-attenders or had very poor attendance. Sometimes, just getting them into the building was a success in itself.”*



On the other hand, five schools explained that their cohorts included pupils with a range of academic abilities making it difficult to design an appropriate intervention.

For example:

- *“For a literacy or vocabulary intervention, there should be an assessment of need based on the literacy requirements of that child”*
- *“These students have an aversion to being pulled out of lessons and are all vastly different in terms of ability. I feel it would have been more beneficial to pull students with a similar level of ability”*
- *“The presumption that they will be academically homogenous is a false one.”*

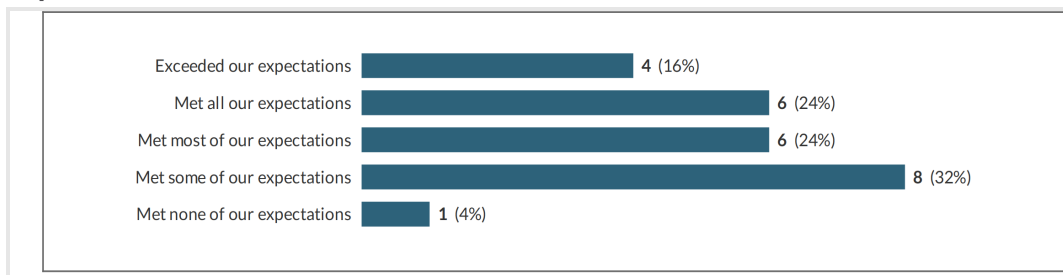
One school also commented that the length of the programme was too short to have an impact. At another school the large size of the cohort was highlighted as a significant challenge:

“There were so many at the start. They were in different classes. The implementation was affected by this which made the whole thing too time consuming and ineffective. The number of staff involved was a massive issue and made much harder than it could have been. We needed to start off smaller and build on successes. I felt under pressure to do too much and in the end this just did not work.”

Overall experience

Schools were also asked to rate the extent to which WfA met their expectations. Ten of the 25 schools indicated that their participation in the WfA programme had met or exceeded all their expectations, while six said that it had met most of their expectations. However, eight schools indicated that WfA had only met some of their expectations and one school said it had not met any of them.

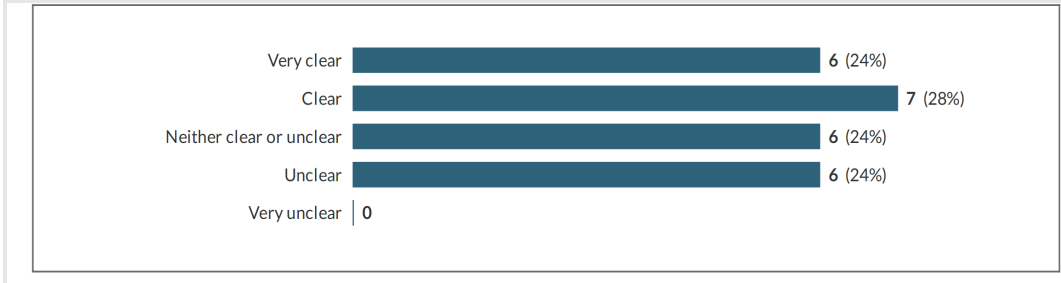
Figure 11. Endline staff survey responses to question on extent to which WfA met their expectations



Schools were also asked to rate the extent to which the expectations placed on schools were made clear at the outset. Thirteen schools indicated that these were made clear or very clear at the outset. Six schools did not express a firm view (selecting neither clear nor unclear), while six schools said expectations had been unclear.

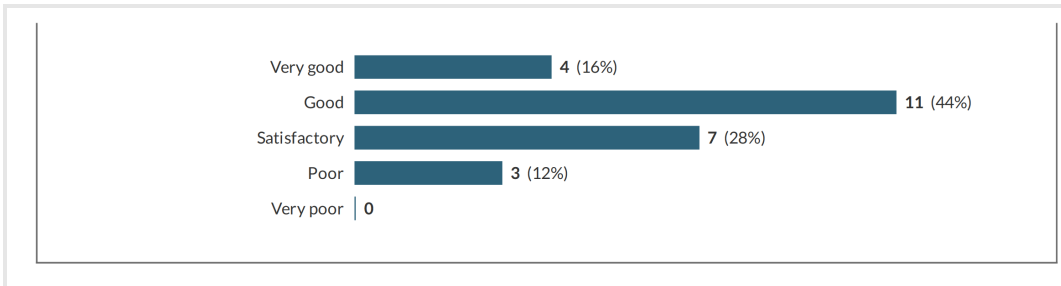


Figure 12. Endline staff survey responses to question on extent to which programme requirements were clear at the outset



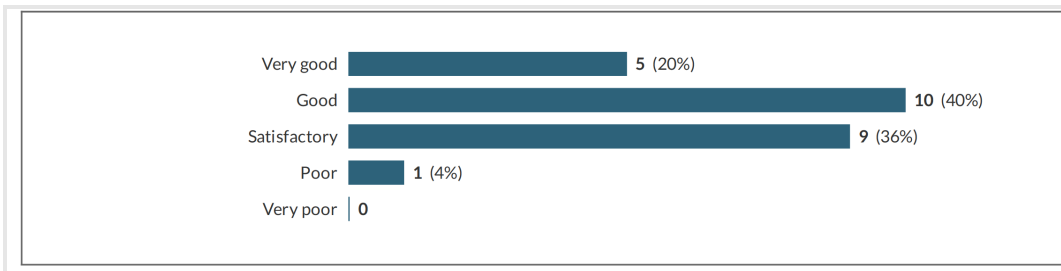
In relation to their experience of accessing and understanding data and evidence on which to base their programme, 15 of the 25 schools rated this aspect of WfA as either good or very good. Seven schools rated this as satisfactory, and three thought this was poor.

Figure 13. Endline staff survey responses to question on accessing and understanding data and evidence



When asked about the overall adequacy of the training provided as part of WfA, 15 schools rated this as good or very good. Nine schools rated this as satisfactory and one as poor.

Figure 14. Endline staff survey responses to question on adequacy of training provided



Further comments

Eleven schools provided further comments at the end of the survey, nine of which were very positive about the programme.

For example:

- *“It has provided a focus around which we could discuss and generate ideas for improving what we offer our students”*
- *“Has been really useful and looking forward to continuing next year with a wider cohort of students”*



- *“A worthwhile experience that has led to some really positive outcomes for our students”*
- *“Effective strategies have been identified through the project we have undertaken, and the involvement in the Words for All has directed the school’s next steps, as well as upskill staff.”*

Two other schools were also positive about the programme but expressed regrets regarding the constraints around implementation owing to COVID-19.

Only one school commented negatively in the further comments:

“I joined the Words for All project after we had signed up and someone else had attended the first seminar and decided it was too much for them in their role. I think I was under the wrong impression as to what the project was about, so I think I found it an additional pressure rather than a fun project. Had I been involved from the start maybe I would have felt differently.”

Interviews with local authorities (June–July 2021)

The original trial evaluation protocol outlined plans for case studies of ten schools. This was to involve interviews with local authority virtual head teachers, social workers and educational psychologists working with pupils participating in school intervention programmes. However, as the trial got under way it became apparent that these professionals were not involved, and in most cases were not even notified about the school programmes. This aspect of the evaluation was therefore replaced with interviews with ten local authority managers and one academy trust manager, all of whom had some awareness of WfA, even if they may not have been closely involved themselves. All interviews were recorded and transcribed in full. These findings summarise six key themes which emerged in these interviews.

How local authorities became aware of WfA

They had been approached by Whole Education during the summer of 2020 to find out if they were interested in taking part in the project. Whole Education was looking for four to six secondary schools in each authority, particularly any that had previously led literacy work and/or had senior colleagues with an interest in “closing the vocab and reading gaps”. In addition, schools had to have the capacity to dedicate three staff (referred to as a “triad”) to this work and be prepared to cascade learning across the authority in subsequent years.⁴

Although some of those interviewed had met with representatives of Whole Education towards the end of the 2021 summer term, there had not been regular communication or updates from Whole Education. With a few exceptions, there was a very low awareness across this group of how the project had progressed and even of the schools that had engaged:

“It was a schools’ project, that’s how I saw it, so basically my role was to make the connection between them and the schools and get the sign-up ... I had assumed we’d be kept up to date with that, that was the impression I was

⁴ It should be noted that the evaluators were not aware of this requirement.



given, but I haven't even had information to say that only one school has been able to continue.” (LA 1)

Although most of those interviewed thought the authority had been approached directly by Whole Education, a few had found out about the project from WWCS's website or because they were exploring possible participation in other WWCS projects. One authority was already working with Whole Education on other projects and had been attracted to WfA because of the link with social care and vulnerability. Not surprisingly, staff working in virtual schools were particularly concerned about Looked After children.

There were several reasons why the project had seemed attractive; not least was the fact that there was very little financial commitment on the part of schools beyond providing cover. All the authorities were concerned about the literacy levels of some pupils and literacy was usually an identified educational priority:

“... the whole thing around vocabulary acquisition and the link between communication and social emotional mental health needed things that are really close to my heart in the job that I do. I was formerly a literacy consultant and school improvement officer, so it's something that has been a lifelong thing.” (LA 2)

The CPD offer was also very attractive to local authorities and, it was hoped, to schools as well.

The way the project was designed implied that there would be a close relationship between the education services and CSC in the local authorities, not least to be able to access the data on children who had a social worker currently or during the previous six years, as well as to alert the social workers of children in the project to their participation. In the event this had not happened to any great extent. A few of those interviewed referred to senior managers in CSC probably being aware of the project and in one authority a data analyst had supported the identification of children that fitted the criteria for participation. In other authorities CSC had said it would not be possible to provide that data. Similarly, there had also been the suggestion that educational psychologists might have some input, but this had not happened in any authority.

Lack of clarity over the programme

The programme was explained as having an evidence base, but it was also stated that it was down to schools to decide the approaches to adopt so they fitted with their existing improvement plans. They understood that support would be given to schools to design their own interventions, but they were unclear how this had happened, if at all. With hindsight most of those in authorities thought that a more prescriptive approach would have been preferable to give schools greater clarity on how to proceed. However, there were a few dissenting voices who thought it was a strength that schools were not told exactly what they had to do.

The original invitation had stated that the local authority lead and contact would be informed, and learning/progress shared and as a result there was some surprise that so little had been asked of them. Although schools had made up their own minds over whether to participate there were



those in local authorities who thought they would have been able to provide the support and encouragement around that decision if they had been able to work in partnership with Whole Education. Others were happy to arrange for schools to be aware of the launch and first hub event and then to take a back seat.

Identifying the schools and pupils

Local authorities have a limited number of staff to fulfil their statutory and strategic responsibilities. It was mainly virtual school staff and school improvement teams that were involved, and they were anxious to point out that their role was one of an influencer and they could only suggest involvement, not direct:

“... there’s a team of five of us, that’s the whole of the education team. Where I know previously local authorities had almost a staff of a medium-sized high school as school improvement teams.” (Virtual head, LA 11)

In most cases the local authorities offered an open invitation to schools to take part, or they targeted schools with a higher-than-average number of Looked After children and where they assumed there may also be higher numbers of children who were the subject of Child Protection Plans. There was only one authority that adopted a more systematic approach to identifying schools with high numbers of such children and where there were significant learning needs:

“I looked at children with specific learning difficulties as well, so I looked at the correlation between high specific learning difficulties – i.e. dyslexia – and high speech, language and communication needs and came up with two schools. I also contacted our social care data team, and I requested a list of children and young people in care in all of our high schools, those on child protection and child in need plans. So, I took out the top seven schools, and then I highlighted four of those which had the highest number of children in care, child protection, child in need, which is what I’d been asked for. I wasn’t aware that it was six years, I didn’t actually find that until Whole Education came back to me and asked me to go back through the whole activity again. I said I was really sorry, but I just couldn’t do it, I can’t begin to tell you how large the job is that I do anyway. And it would have been a struggle for social care to have pulled up children who’d ever had a social worker, which was the criteria, in the past six years.” (LA 2)

But even this person saw the authority’s role as facilitating the link with the most appropriate schools:

“... so over and above attending part of the launch event so that I understood what it was going to offer, I’ve not had any other involvement there.” (LA 2)

Others found it more difficult and even impossible to help schools pull together the data on pupils. Virtual head teachers spoke about having an overview of the current cohort of Looked After children but did not have historic data. Some had entered discussions with CSC but, apart from the one instance outlined above, had been told that it was not possible to provide the level



of data required against school information for the past six years for children living in the authorities' boundaries, let alone those who were not their responsibility.

Some of those interviewed had also heard that some schools had faced difficulties obtaining the data that was held internally. Although safeguarding leads would know the children involved with children's social care, that knowledge is held very securely by schools, and the WfA project leader would not have access. This was then compounded by remote working where transferring data securely became a problem. In at least one school most of the children who were in the targeted cohort were placed in alternative provision and so not able to be included. Overall, local authority staff thought that only a proportion of eligible children will have been identified.

It was also noted that some of the schools that the local authorities thought would have benefited from participating had very few eligible children because they had been judged by the inspection body, Office for Standards in Education, Children's Services and Skills (Ofsted), to be inadequate and Looked After children could not be placed in such schools.

Among those interviewed there was a recognition that Looked After children and others who had been involved with CSC *may* be at higher risk of educational disadvantage and, as a result, there was some sympathy for the emphasis on children who had been involved with CSC. However, about half of those interviewed did not agree with the timescale "of involvement in past six years" or with the necessary linkage between CSC involvement and an educational deficit:

"I think it's a little bit concerning if we were to assume that all young people who had had experience of social care would somehow be academically disadvantaged, and I think within education it's being really mindful of those broad assumptions." (LA 5)

"I was very uncomfortable about the focus on children who had a social worker over past six years. We have many children who need an intervention like this but have not had any contact with social workers. It was a poor call." (LA 8)

Such views usually went hand in hand with concerns about how children would feel about being identified in this way:

"I suppose one of the things that was key for me from a sensitivity perspective would be as to whether the young people knew that that was why they had been selected for a project. Thankfully within the intervention groups you were allowed to include others, so as long as it had the identified pupils, you could then invite pupils that were not on the list at all but just would benefit from the intervention. But young people are very astute, and I do wonder if they might have looked around and thought, hang on, maybe I know why we're all in this group." (LA 4)

"I have had some feedback from a school that pupils have not engaged well. I do wonder if part of the problem is that they just did not want this type of exposure and I don't blame them." (LA 8)



Content

Most of those interviewed said they were not able to comment on the details of the content of the programme. The only exceptions were two experienced English teachers who expressed reservations about the books that had been suggested as suitable for the cohorts involved. Both acknowledged that this did come down to personal preference and experience of how pupils in different year groups had responded to texts:

“I can see that some of those have been chosen potentially because they’re quite wordy texts, almost we feel you should have read this. There were some what I call popular novels on there, which I thought was great. I think there were occasional ones where I thought, I’m not sure I would have read that to that year group when I had been a teacher ... and thinking that maybe it felt a little bit too mature. Occasionally I felt it the other way as well. I felt that was a bit of a young ... but I think it’s very subjective in terms of the choices.” (LA 5)

“I think some were just odd and I could not understand why they would have been suggested. Take Stories and Poems for Extremely Intelligent Children of All Ages. Why? If they think you spark an enthusiasm for reading by giving children that, I despair! You can fire the imagination in so many ways – that is not one of them. It was one book, but it is worrying.” (LA 9)

Why schools dropped out

The backdrop to many schools not continuing to engage was the pressures under which they had operated during the 2020/21 school year, including reduced capacity:

“And it’s a waste of money; I don’t know how much funding it would have saved by stopping it and calling it off, but even if it was only a little bit, to breathe again and run it another day it would have been worth it, because nothing was going to happen during the pandemic.” (LA 2)

“I can only imagine how little bandwidth people in schools had for additional activities in the last 18 months. So people just turned inwards and thought, right, we’ve just got to try to keep this school open. So being remote to start with, because I think COVID kicked off just at the time that they were trying to launch, meant that they didn’t have any live face-to-face contact, they didn’t actually meet the people who they were trying to engage with, and getting commitment from people like that I think was really difficult.” (LA 11)

Other factors were suggested. In some cases, the combination of the workload associated with the project, particularly the identification of intervention and control groups and additional testing, had combined with the pressures associated with the pandemic to lead to the decision. Schools had chosen to or were forced to focus on their immediate priorities, and this was described as “an unnecessary distraction and burden”. At least one school had fed back that:



“... they found that very difficult to cope with the idea that you would identify a group of students and then not intervene for some of them, they really struggled with that. ... they felt that they would be putting too much work into a tiny number of students when we were in the middle of COVID-19.” (LA 3)

In other instances, local authorities had been told that the schools had a literacy improvement programme in place and that adopting this project would not be helpful. Others referred to a reluctance on the part of schools to take part in an initiative across mixed year groups and had fed back that they would have preferred to have confined it to Years 7 and 8 or Years 9 and 10.

Views on reshaped intervention

Because there had been very limited communication between local authorities and those schools that were taking part in the project most of those who were interviewed were unaware that the intervention was now taking place over a minimum of six weeks. The reactions to this news were perhaps not surprising. They expected it to have a very limited impact for several reasons:

1. While there may be some improvements to how individual teachers reframe their lessons and understanding of vocabulary, lasting improvements require a long-term literacy strategy. A part of this is the school's approach to reading. If the project had taken place over the original timescale there had been the hope of achieving sustainable improvements which were unlikely to occur over such a short period
2. Some of the local authority staff interviewed had received positive reports of the progress recorded in schools. While this was welcomed, the fact that summer holidays followed soon after the end of the intervention would probably mean that even if improvements had been achieved, they were unlikely to survive the break. Nevertheless, there was the opportunity to build on what had been learned to inform planning for the next school year.

Overall, most of those seen in local authorities thought it was a waste of resources to have run it during the 2020/21 academic year. They thought the project should have been halted and restarted in September 2021 – there was just too much going on for them to be able to commit:

“I do think in a way it's a shame that it wasn't just paused. They should have said 'we know you're interested, we've got this funding, let's just think that 2020/21 has been too severely interrupted and let's launch again. I know practically that there are all sorts of implications for that, but I think from a school's perspective ... , even after the schools reopened towards Easter, so many of them have had interruptions just because of bubbles of pupils needing to self-isolate, teachers themselves, and I think ... what I would be worried about would be whether the experience of trying to do a project in such a challenging year, if that might cause any hesitancy for schools moving forward, actually the vocabulary project was too hard.” (LA 5)

Case study interviews (November 2021)

The original plan for this evaluation was for the researchers to visit ten schools to interview key staff involved in the project and undertake focus group discussions with participating pupils in



May and June 2021. This had to be pushed back until the autumn 2021 term after the implementation of school intervention programmes was delayed. The 25 schools that had completed the post-intervention survey were invited to participate in the case studies and 10 schools agreed. In the context of COVID-19 it was not possible for the research team to visit schools in person and interviews took place online. Furthermore, schools were only able to commit to one online interview and the focus group discussions with pupils were not possible. In the end nine interviews were carried out during November 2021 as one school did not show up for a scheduled interview on two occasions. All the interviews were recorded and transcribed in full.

Background on the nine schools

Seven of the nine interviewees said their schools had been alerted to the project by an email from their local authority or academy trust which resulted in a member of their senior management teams wanting their schools involved. The other two schools were already working with Whole Education on other projects and were invited to take part. The key members of staff who were usually involved in designing and implementing the response were based in the English Department, either with responsibility for literacy or working alongside colleagues with that responsibility. In some cases, subject teachers from other subject areas were also involved and, in a few cases, teaching support workers and library staff. There was only one school where just one member of staff – the literacy coordinator – was involved in running the project on their own.

In four of the schools the intervention cohort was very small (ten or fewer pupils) and much smaller than the cohort allocated by WWCS. In the remaining 5 schools the intervention cohort ranged from 13 to 60 pupils so was broadly in line with the allocated cohort.

Clarity over expectations

Intervention group

While most of the nine schools said they were clear about the top-level expectations it was evident from the discussions that there were some misunderstandings about the pupil group and the tests associated with the impact evaluation. One school had added some more pupils to the intervention group in the hope of avoiding pupils in the group thinking they had been targeted and, as far as it was possible to determine, these additional pupils had not been excluded from testing.

Two other schools had also augmented their samples with pupils that were outside the specified group (i.e. any pupil that had a social worker in past six years). Although technically coming in this group one school included pupils with an education, health and care plan (EHCP) pupils as, during COVID-19, they had been allocated a social worker. It was noted by the interviewee that it was pupils in the EHCP group that had made most progress. Another school had included all pupil premium pupils in the project.

The interviewee said:



“I don’t ever remember being told specifically they had to have had a social worker in the last six years, to be honest.”

Elsewhere one interviewee said they had been told by Whole Education that schools were typically submitting the names of around 100 pupils. They had initially thought that they could include all free school meals pupils, but when they found out this was not the case, they ended up with 17 altogether, nine in the intervention and eight in the control group. This final number had come as a surprise and, as a result, they wondered if other schools had submitted data on a larger group. There was also a school that decided to involve only one year group – Year 9 – to make delivering the intervention more manageable; they wanted to do the same thing with all of the pupils involved and map it onto what they were doing in their English lessons. This school had asked the local authority to identify the sample, but data were not provided so they had to work with data available in the school.

Nature of the intervention

Three schools planned to use a Pre-teach Vocabulary (Miller and Veatch, 2011) intervention. In one case staff in the science department used the Frayer model (Frayer et al., 1969) which was implemented as planned. Another ‘pre-teach’ school had intended to use the approach in geography and in science, but, as the project’s timescales became shorter, they decided to use it only in geography. They then moved away from the planned approach to one where they read a text and discussed it with the group, focusing on supporting them to understand the vocabulary being used in lessons. The third school only introduced the intervention into Year 9 where it was used in film studies as this was not a GCSE subject so caused less disruption.

The other six schools all used distinct interventions. In one it consisted of discussion groups and reading sessions. The former was supported by a programme, written by the pastoral manager, to encourage pupils to listen to other people’s opinion, and then to agree or rebut the argument. It was not intended to benefit literacy specifically but rather to improve pupils’ confidence and social skills. In another school all teachers were encouraged to write about an aspect of their life and these accounts were read by the pupils who each chose a teacher and wrote a letter in which they explored the account in more detail. The pupils also wrote an account of an episode in their lives and the lead member of staff for the project was bringing all the pieces together in an anthology.

A third school used a comprehension exercise with pupils which involved them reading articles on subjects not necessarily covered in the school curriculum and then answering a series of questions on them. Another school used the Read Aloud approach (Harvey and Goudvis, 2000; Quigley, 2020). The third and fourth schools used activities designed to enhance vocabularies with the latter also using a project to encourage parents and teachers to make sure subtitles were on the screen when pupils were watching a TV programme or video. The idea of “same language subtitling” (SLS) is that pupils hear and read at the same time (d’Ydewalle et al., 1991; Parkhill and Johnson, 2009). However, the SEN department in that school picked up on the approach and sent a general request out to all parents of pupils they worked with to switch the subtitle option on.



In all nine schools the intervention programme could not begin until after the Easter holidays (mid-April 2021) because of the COVID-19 restrictions. One school had been ready to start in January as originally scheduled, but elsewhere schools had not finalised their plans before the Christmas break and the delay allowed them to continue planning. However, in three schools the intervention did not begin until mid-May or in one case until June 2021.

The post-intervention survey results reported earlier showed variation in the length of intervention programmes across schools and that in some schools it had lasted for less than the minimum six-week period that had been specified. Among the nine schools interviewed the programme had lasted for six or more weeks at six schools. However, in two schools it had only lasted for five weeks and at one for only four weeks. Interviewees across all schools highlighted the difficulties of planning and implementing the programme in the context of COVID-19 and recognised the limitations of such a short intervention. Some schools commented that the trial should have been delayed.

Several schools would have preferred to have had a defined project to implement, and this was the understanding on which they had signed up. One of these schools wondered why they were being expected to take part in a RCT when they considered the number of variables, the short implementation timescale and the shifting context of COVID-19 undermining any attempt to establish a valid methodological approach.

Notifying parents

All nine schools had informed parents or carers that their children were taking part in the project although it was not clear how many had gone into details of how their children came to be in the sample. One school did refer to a using template provided by Whole Education that did refer to pupils' involvement with children's social care services; however, another school seemed to have deliberately avoided making the selection criterion explicit:

"I just said we've got a range of students across a range of years of different reading abilities. I didn't go down the whole social care route too much because, for some of our parents, it might put them off massively, and I just said we're doing some educational research with a university, and they seemed to like that."

Two parents from one school refused permission for their children to be involved: one did not think their child needed it and another did not want the child to be withdrawn from class. Another school had a complaint from a parent whose child was in the control group. It was nearly a year since they had been informed about the project and the WfA contact had found it very difficult to say to the parent "your ... child is in the control group, your very needy SEND child was in the control group". In that event, parents were told that the project was ongoing, as the head teacher was intending to fund it for another year, and as a result their child would have the opportunity to participate:



“So, we were able to put that sort of spin on it, but had we not been repeating it, it would have felt really very, very awkward indeed to have that control group. I know why we have to have the control group, so you can measure the impact, however we’re talking about students in a particular group, and they could have had some extra input in the project.”

Testing

New Group Reading Test (NGRT)

The intention was for pupils in the intervention and control groups to take the NGRT at T1 (before the intervention) and at T2 (after the intervention). Most of the schools were familiar with the NGRT but there were still comments about links to the site not always working and one interviewee described it as a very clunky system to negotiate.

Eight of the nine schools had used the NGRT, at least at T1. In the ninth school the interviewee said they were not aware they had to do this and when they found out it was too late to do it at T1, so they had been told by Whole Education to monitor the pupils’ progress.

Generally administering the NGRT at T1 had not been a problem apart from schools having to work around bubbles and deal with pupil and staff absences. However, in four schools not all the relevant pupils had done the test, either because some pupils’ learning difficulties made it impossible or because they had been absent for periods of time. At the time of the interviews (mid to end November 2021) only two of the eight schools using the NGRT had completed T2 testing.

Strength and difficulties questionnaire (SDQ)

The SDQ was meant to be used at T1 and T2 to assess any changes in the emotional wellbeing of pupils in both groups. Only one interviewee said it had been done at T1 and T2. In four of the nine schools those who were interviewed said they were not aware that it had to be completed. In the other four schools it had either not been attempted or a link had been sent to pupils but those interviewed were not aware to what extent it had been completed.

Challenges

Schools reported a range of challenges over the implementation of the project, although the difficulties associated with running it during the pandemic were universally highlighted. COVID-19 meant schools had been providing teaching online between January and early March 2021, so the planned start for January 2021 was not feasible. Even when they returned many staff and pupils were absent for periods of time with the virus. Government advice to keep children in groups or “bubbles” often made it more difficult to deliver the interventions as planned. This was further complicated when pupils were already in GCSE option groups. In one school this meant that the original sample of 50 fell to just 20 across the intervention and control groups. Even when schools opened, if pupils in a bubble became infected everyone in that bubble was sent home. There were attempts to make it possible to access the intervention remotely but where this had happened the feedback was that it had not worked:

“... because so many children, as during the lockdown, it was hit and miss whether they would engage, whether they would log on, whether the parents



had capacity to support them to log on, there were so many variable factors which impacted on this.”

Consequently, there were several references to how few pupils in some schools had received the “full” intervention. However, this was not a complete surprise as even in “normal” times staff said that many of the pupils in the sample were among those with the least consistent attendance, most likely to move and the most difficult to track.

Although senior managers and other colleagues were said to be supportive, the additional pressures associated with maintaining schools during the pandemic, compounded by the fact that there was no additional off-timetable allowance, were said to have made it difficult to arrange planning meetings in schools and attend the Whole Education webinars and hub meetings.

Unrelated to the pandemic, one school experienced a significant delay in receiving information on the randomisation process which had shortened the time for the intervention. Elsewhere there were references to the problems associated with maintaining the divide between the intervention and control groups. The example given above where the intervention had also been adopted by the SEN department was just one instance where fidelity to the methodology may not have been maintained. Other interviewees said they did not think all those that were eligible had been included (that is, any pupil who had had a social worker in the previous six years).

Several interviewees had felt constrained by the requirement to work with this group of pupils when they believed either that whole classes would have benefited from the approach or that it meant some gifted and/or able readers were included while pupils who were in need of additional literacy support were excluded. Some of those who were involved from a very early stage (by September 2019) had been surprised that they could choose their own intervention when they had been led to believe that they were taking part in a more defined project accompanied by a RCT. These views are best summed up by this head of literacy:

“When we were given this empty project plan and just said, well, it’s all to do with vocabulary and gaps in vocabulary, you decide what you’re going to do, then funding came into it obviously, because we didn’t have any money, we didn’t ... there’s no budget for this, and we’re a local authority school so there definitely is no budget. We had to think about what we already had in place because this was a time when lots of people were off, lots of cover was taking place and we had to make sure that our own workload was protected as well, so we just didn’t have time to write massive schemes to start from scratch, which is why we went for [names intervention].”

Support provided by Whole Education

Most of those who were interviewed were satisfied with the support received from the staff of Whole Education before the intervention commenced, but several commented on how this had fallen away during the intervention. Furthermore, one interviewee was left wondering what Whole Education had contributed over and above providing the webinars and inter-school contacts as it had been up to schools to determine and plan the intervention.



Not all those who were interviewed had attended the webinars, either because they had joined the project later than others or they did not have non-teaching time.⁵ Those that had attended judged them to have been of a high standard, although more relevant to improving literacy in schools in general rather than to their projects. Similarly, they had not all attended the hub meetings arranged for participating schools but those that had generally found it useful to hear what other schools were doing. However, one interviewee thought that Whole Education should have done more to challenge what seemed to them to be “mundane, low-stakes” projects to maximise the effectiveness of the project. Another interviewee would have found a regular newsletter helpful.

All schools were allocated a mentor. Most interviewees found the contact very helpful, although in some cases the pressures discussed above led them to reflect that they had not used the arrangement to its maximum and would have welcomed clearer direction over the purpose and expectation of the project. However, two interviewees were not able to comment as they had not been able to establish contact with their mentors.

Pupil engagement and achievement

On reflection all those who were interviewed thought that the engagement of pupils had improved after some initial difficulties. These difficulties were associated with pupils not wanting to be identified as in any way different and being suspicious of what was happening.

In addition to providing reassurance, schools had addressed this by holding launch events and providing snacks during the session which, in most cases, was thought to have increased trust in, and even enjoyment of, the project. The exception to this was one school’s failure to engage pupils who were in an alternative provision that was attached to the school.

All schools thought that pupils had benefited from taking part in the project alongside various references to how some pupils had grown in confidence. However, as far as school staff were concerned, the reality was that it was a very short intervention delivered during a time of particular difficulty, so any expectations had to be seen within those constraints.

Curriculum development

Six of the nine interviewees referenced how the project had had an impact on curriculum development in the schools, one even described it as a “spur to future activity”. In most cases the developments related to an increased focus on literacy across the school and specifically the teaching of vocabulary. In this context the following comments are worth noting in relation to the importance of teaching academic vocabulary to complement subject teaching:

“Teachers have recently had CPD from me, talking, breaking down the different methods and explicitly teaching vocabulary, they’ve got that in their

⁵ Recordings of the webinars were made available for later viewing.



arsenal, they know how they can implement it in their lessons, so I do think that has been fruitful.” (Literacy coordinator)

“As part of the disciplinary literacy policy there is CPD introducing the policy and then at the start of this 2021/22 year had a whole staff CPD where Science teacher talked about what the Words for All project was, what they had done, whom they had with, the impact of that and how that was shaping what she was doing in science. They want every subject area to be looking at Pre-teach Vocabulary. Staff have to identify the significant vocabulary for a unit.” (Head of English)

Cost analysis

The programme costs of the three scenarios modelled range from £1,995 per school to £2,660 per school. This results in an average cost per pupil beneficiary of £81 to £94. These results are presented in more detail below.

Unit costs used to value teacher triad time are shown in Table 5. In summary, annual costs of employment are estimated for a lead staff member (assumed to be employed at spine point 9 on the “lead practitioner” teaching scale) and for classroom teachers employed at spine point M4 on the national pay scale for teachers. Direct and indirect overheads and other employment “on-costs” are approximated from published unit costs for local authority employed social workers (using the ratio of social worker salary to corresponding overheads and “on-costs” and applying this to the teacher salaries as a rough approximate). Total annual costs of employment are then divided by an estimate of annual teacher workload (measured in hours) to obtain an estimate of the opportunity cost per hour of time spent on specified activities. All costs are expressed at 2021/22 values.

Table 5. Unit costs (2021/22 values)

	Cost per hour	Notes and assumptions
Lead staff member	£47.48	Based on annual pre-tax salary for “lead practitioner” teacher at spine point 9 (NASWUT, 2022). Salary on-costs plus direct and indirect teacher overheads are added and assumed to be 86% of pre-tax salary based on the ratio of salary: on-costs/overheads for other LA employees (social workers) (Jones & Burns, 2021). Cost per hour is estimated by dividing total annual cost of employment by average annual teaching hours. Estimated working hours per week assumed to be 51.6 for a ‘middle leader’ based on Department for Education Teacher Workload Survey (Walker, Worth & Van den Brande, 2019). 39 working weeks assumed over one school year.



Other triad teachers	£31.53	Assuming an annual pre-tax salary for a classroom teacher at spine point M4. On-costs and overheads as per above. Teaching hours assumed to be 48.1 per week (Walker, Worth & Van den Brande, 2019) over a 39-week school year.
----------------------	--------	---

Table 6 presents the estimates of the cost of the lead staff member engaging with provider training and support.

Table 6. Cost of lead staff member time allocated to provider training and support (per school; applicable to all project scenarios)

Activity	Total teaching time	Incremental cost
Enrolment conversations/activities	2 hours	£96
Pre-thinking event	3 hours	£143
Launch event	4 hours	£191
Hub events	9 hours	£430
Expert input webinars	12 hours	£573
Support conversations/visits	3 hours	£143
Report writing	2 hours	£96
Impact event	6 hours	£287
Total hours and cost	41 hours	£1,959

Table 7 presents estimates of the cost of specified project activities per school under each project scenario.

Table 7. Cost of teacher time for alternative project scenarios (per school)

	Total teaching hours allocated to project activity	Cost of teacher time	Notes and assumptions
<i>Scenario 1: whole-class intervention delivered in lessons. Duration = 6 weeks</i>			
Programme development time	9	£332	1 x lead practitioner (spine point 9) and 2 x other teaching staff (spine point M4) for 3 hours each.
Material prep time	9	£332	10 mins of prep time per lesson across 18 lessons x 3 teachers (grades as per above).
Direct contact time with pupils	54	£1,995	1 x lead practitioner and 2 x other teaching staff (grades as per above).
Total	72	£2,660	
<i>Scenario 2: small class/1:1 intervention carried out after school hours/during tutor time. Duration = 6 weeks</i>			



	Total teaching hours allocated to project activity	Cost of teacher time	Notes and assumptions
Programme development time	4	£191	1 x lead practitioner at spine point 9.
Staff briefing/training with colleagues	12	£411	Lead practitioner trains 5 teaching staff (5 x spine point M4).
Material prep time	9	£261	1 x lead practitioner and 5 x other teaching staff (grades as per above).
Direct contact time with pupils	54	£1,565	30 mins x 18 sessions over 6 weeks for 6 members of staff (grades as per above).
Total	79	£2,428	
<i>Scenario 3: out-of-school project with parental engagement delivered at home – duration= 6 weeks</i>			
Programme development time	30	£1,109	Longer lead-in time for programme development involving 3 teachers (2 x spine point M4, 1 x lead practitioner at spine point 9) for 10 hours each.
Briefing sessions with parents	6	£222	1 x lead practitioner and 2 x other teaching staff each delivering 2-hour briefing session with parents (staff grades as per above).
Creation of materials for home use	18	£665	1 x lead practitioner and 2 x other teaching staff each inputting 6 hours to material development (staff grades as per above).
Total	54	£1,995	

Irrespective of the type of project initiated within schools, it is estimated that 41 hours of lead staff member time would typically be required for attendance at events focused on provider training and support. This would equate to a cost of £1,959 per school enrolled into the programme. Total teaching hours allocated to project-based work within schools would vary between 79 hours (£2,428) for small group/1:1 work and 54 hours (£1,995) for a home-based intervention over a six-week period. The lower estimated teacher hours required for the home-based project reflects the reduced time allocated by teachers to face-face work with pupils, with an increased emphasis on parental time invested in working directly with children.

Table 8 presents cost per pupil beneficiary estimates.



Table 8. Cost per pupil beneficiary

	Cost (lower and upper estimates)
Scenario 1 (lesson-based)	£94 (£15–£577)
Scenario 2 (small group/1:1 after school)	£90 (£14–£548)
Scenario 3 (home-based)	£81 (£13–£494)



Limitations

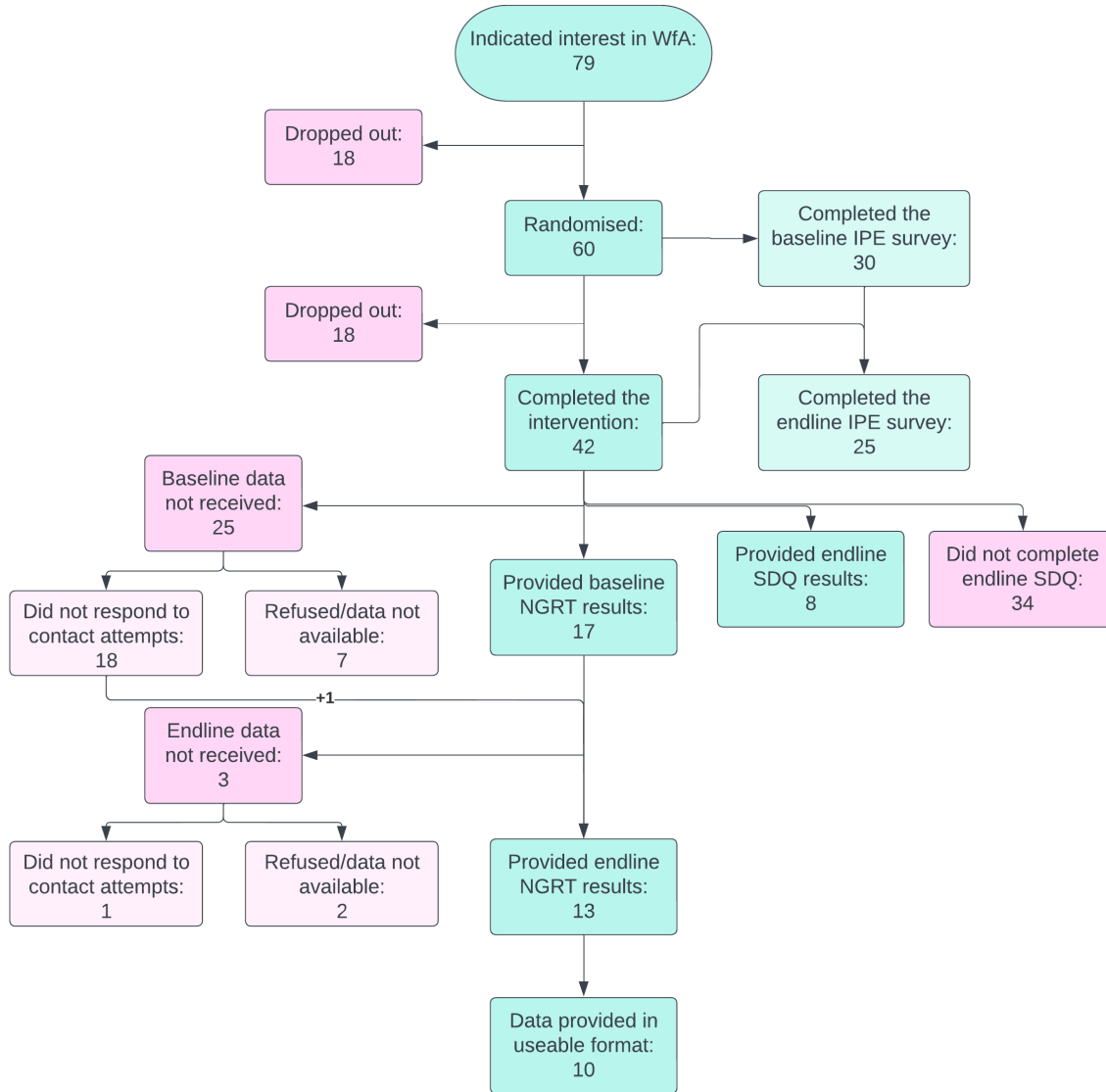
It should be noted that this trial took place during one of the peaks of the COVID-19 pandemic, resulting in significant difficulties for both programme implementation and evaluation data collection.

The final sample size was substantially smaller than anticipated, resulting in an under-powered RCT. When designed, it was anticipated that approximately 2,400 pupils would participate in the trial. This sample would have allowed the RCT to detect a small-to-medium effect size on the primary outcome. However, a number of schools dropped out over the course of the programme. Based on feedback received through the IPE, this is most likely due to schools struggling to cope with restrictions brought in due to COVID-19. At the end of the 2020/21 school year, Whole Education reported that 42 schools had participated in the programme, still representing over 2,000 pupils. However, when the research team contacted schools to organise data collection, only 17 schools provided any data, despite repeated contact from both the research team and Whole Education.

The flowchart below shows the level of attrition across the various stages of the project.



Figure 15. Level of attrition across the various stages of the project



Consequently, the final sample size was 358 pupils, after excluding duplicated observations. Out of those 358 pupils, 221 were complete cases, 62 only had endline data and 75 only had baseline data. The 221 complete cases represent only 10 schools.

The final treatment and control groups available for analysis were not balanced. From the total participant sample, 80% of pupils were in Years 7 to 9 and 48% of pupils were female. Figures 17 and 18 present the sample’s distribution by year and sex.



Figure 16. Distribution of pupils by year of study (of final sample)

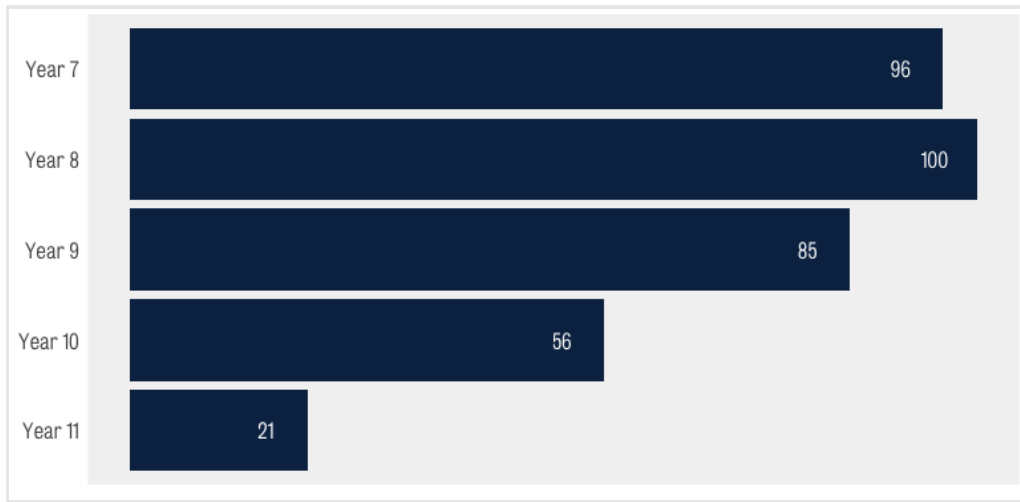
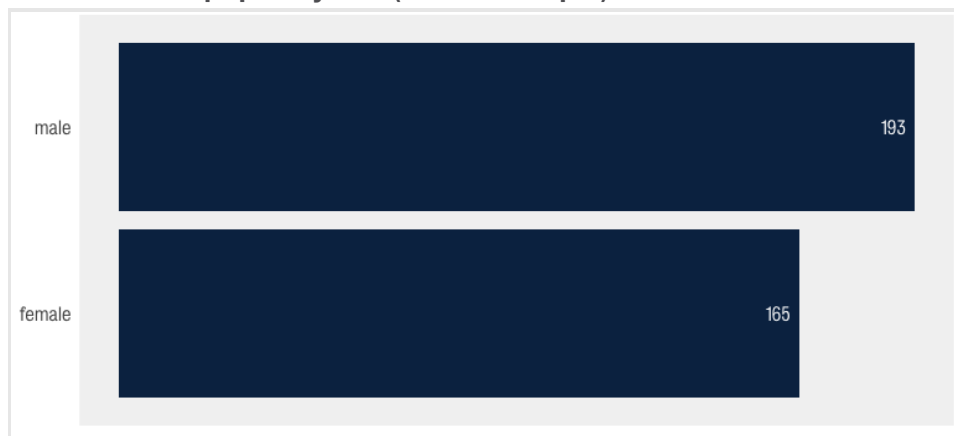


Figure 17. Distribution of pupils by sex (of final sample)



For an RCT to be successful, treatment and control groups should be balanced. This means that on average, there are no observable differences across groups. Based on the available data it was only possible to balance check for sex, and this analysis found that the treatment and control groups were not balanced for this variable.

For the final sample, there was an approximately 10% difference in the allocation of male and female pupils to treatment and control at baseline. Considering a standard tolerance threshold of 5%, these groups are therefore unbalanced at baseline. Therefore, the results in this report should be treated with caution, given that the analysis may be comparing treatment and control groups that were not similar on average at starting point.



Table 9. Sex distribution in baseline treatment and control groups (of final sample)

	Female	Male
Control	0.52	0.48
Treatment	0.41	0.59

The sample for the complete data set (both baseline and endline NGRT) is also unbalanced. At baseline, there was a 10% difference between male and female pupils in the treatment and control groups. This suggests that a larger share of male pupils dropped out of the study.

Table 10. Sex distribution in treatment and control groups of complete cases

	Female	Male
Control	0.54	0.46
Treatment	0.44	0.56

The trial has imperfect compliance. Compliance refers to the degree to which participants take up the treatment condition they were assigned to. In this trial, only 79% of the pupils received their original assignment condition. While there are statistical mechanisms to correct for this situation, these mechanisms are also constrained due to the limited sample size. It is not known why some pupils did not receive their allocated condition but, as raised in the IPE, issues with the complexity of separating treatment and control conditions within the same class may have been a factor, as could have been the view raised by teachers and local authority staff that CSC-experienced pupils were not the correct population for this intervention. For example, during data collection one school informed the research team that they had thought it more appropriate to deliver WfA to the whole of Year 7, and therefore did not refer to the randomisation at all when selecting which pupils would receive the intervention.

Concerns about pupils' eligibility threaten the external validity of the results. External validity refers to whether the trial results are generalisable to the wider population of interest. This trial was designed to evaluate the effect of WfA on pupils with CSC experience. In the IPE, however, staff members reported issues with the data lists submitted to identify the pool of eligible pupils in each school. The lack of clarity over how this information was compiled inside schools raises reasonable doubt about whether they accurately reflected the sample of pupils with care experiences. Moreover, staff members referenced barriers to accessing official CSC data from local authorities. One interviewed staff member also highlighted their lack of knowledge around who was in scope for the intervention in the IPE. It is therefore plausible that pupils out of the target group participated in the trial.

Conducting endline testing in October instead of June may have resulted in a reduction in the treatment effect. Some endline testing was conducted after the summer holidays as few schools had completed endline testing before the end of the previous school year. While it is possible that this led to a reduction in any uplift to reading skills caused by WfA, given the small lifts observed through post-holiday testing it is unlikely that pre-holiday testing would have produced different enough results to be statistically significant.



Only around half of the schools that implemented the programme completed both the pre- and post-intervention surveys. Despite significant efforts from the research team, the post-intervention survey was completed by 25 schools, 60% of the 42 schools Whole Education reported had implemented the programme. Survey respondents were self-selecting, meaning that there may be a reason that the other 17 schools did not respond, and they may have provided significantly different responses. Findings from interviews with local authority managers and staff in schools indicate that there was a lack of clarity regarding both the content of the programme itself and the requirements of the evaluation. These may have been additional factors, alongside COVID-19 pressures, influencing schools' decisions to drop out of the trial and for the failure of some participating schools to comply with the conditions of the trial.

Lack of school-based activity data meant that the cost analysis was limited to a scenario-based model. The analysis was based on assumptions about the types of projects that might be undertaken if the programme were to be rolled out more widely in the future. Teacher resource requirements and the associated cost-implications for different types of projects are variable: while home-based work may be the least expensive type of intervention in terms of teacher input, it would potentially shift the resource burden of project delivery from teachers to parents (parental time impacts were not valued as part of this study).

The scenarios are broad characterisations and will not reflect the likely variation in practice that could result in roll-out of the programme and the subsequent variability in resource implications on a school-by-school basis. The scenarios also reflect 'ideal' levels of project resourcing. Actual project-based delivery in the RCT was severely limited by the COVID-19 pandemic, as evidenced in the IPE. The cost analysis also assumes that the resource requirements underpinning each of the scenarios are incremental in nature; in practice this may over-estimate the true additional cost of the programme if schools already invest some teacher resource in vocabulary-based work within the target population of pupils.



Discussion

Discussion of findings

One of the aims of the evaluation was to assess the extent to which fidelity was maintained and, indeed, what constitutes fidelity for the WfA programme. However, during the early stages of this evaluation it became clear that schools were to be given significant autonomy in terms of the content of their intervention programmes. Schools were advised, but not required, to use or adapt one of five evidence-based programmes (later reduced to four) by Whole Education. Moreover, schools were free to follow a 'pick and mix' approach, selecting elements from these programmes or any other they wished to use. School resources allocated to support the implementation of programmes also varied, as did the length of the intervention period. As such there was no evidence of fidelity to a programme, a finding which severely compromises the RCT, as well as the cost analysis.

Furthermore, it was also found that schools did not universally adhere to their allocated treatment and control cohorts or meet all the requirements regarding the collection of outcome data. This, as well as the small number of schools that remained in the trial at the end of the intervention period (July 2021), further compromises the trial.

Overall, the evaluation found no significant impact of WfA on reading skills. However, as the RCT had a very small final sample and significant issues with compliance and data quality this finding has low confidence.

Another aim of this evaluation was to learn about the experiences of staff involved in WfA. Clearly, the experiences of school staff, and the degree to which they were able to engage with the programme, were severely impacted by the COVID-19 pandemic. Most schools were positive about the training and support provided by Whole Education, but some acknowledged that they had struggled to attend many of the hub meetings or expert webinars or to link up with their mentors. However, some schools also reported that their mentors had not been responsive. In terms of their WfA programmes in school, it was generally accepted that the implementation period had been too short to have much impact on pupil achievement, but some schools spoke positively about the impact on pupil engagement, staff development and curriculum development.

There were some interviewees, particularly those representing local authorities, who questioned why the implementation of WfA programmes in schools had not been delayed, given that COVID-19 had severely limited the capacity of schools to develop and implement their intervention programmes. Furthermore, local authority interviewees and some schools also questioned the choice of cohort for this trial. It was not evident that the 'data lists' of eligible pupils had been checked against local authority databases and that left open the possibility that some pupils may have been missed, particularly those living in an authority that is different from the one where the school is based and if their involvement with children's social care was



historic. These issues with eligibility further threaten the findings of the trial, as it is also possible that pupils who were outside the eligibility criteria participated, and therefore the findings may reflect the results of pupils who do not have CSC experience.

Beyond eligibility, staff also felt that, more fundamentally, the assumption that all children in this cohort needed additional academic support should be questioned and it was suggested that the cohort should have been selected on schools' assessment of pupil needs. Concerns were also raised by some about the potential stigmatising effect of taking children out of lessons because of their involvement, or previous involvement, with children's social care.

Implications for future research

Similarly to the findings from this study, the 2014 EEF efficacy trial found a small positive effect on pupils' reading skills, but the finding was not statistically significant. However, the findings from the 2014 trial are more secure than the findings from this trial, due to a substantially larger sample size (final sample of 570) and attrition of only 12% (Styles et al., 2014). The findings of both the 2014 trial and this trial suggest that if another RCT of WfA were to be conducted, it would be important to ensure the trial was adequately powered to detect a potentially small effect.

Further, if WfA is to be evaluated again the success of the evaluation will likely depend on the buy-in of participating schools. In this trial, low buy-in from schools was a key barrier to accessing data. Though the unique difficulties caused by COVID-19 were clearly a factor, it is likely other concerns about the intervention and the research affected schools' willingness to participate. For example, concerns raised by both schools and local authorities about the appropriateness of selecting CSC-experienced young people as the target cohort suggest a poor understanding of the rationale and purpose of the trial. In future evaluations, more extensive pre-trial consultations with stakeholders may increase the likelihood of a successful RCT.



Conclusion

The WfA programme and this evaluation were delivered through an extremely challenging period for schools. Staff reported that their ability to engage with the programme and deliver it to their pupils was severely impacted by COVID-19. Although some staff were positive about their experience of working with Whole Education, others had been unable to engage, and some raised concerns about the lack of support provided. Further, changes to the programme introduced in response to the pandemic, such as shortening the intervention period to six weeks, were perceived as undermining its ability to deliver results. Concerns were also raised about the appropriateness of CSC-experienced pupils as the target cohort for this intervention.

In this context, the impact evaluation did not find any evidence that WfA had any significant effect on pupils' reading skills. While a small upward trend was identified in the average NGRT results of the treatment compared to the control group, it was too small to have confidence that it was not due to chance. Substantial difficulties with data collection and data quality resulted in a small sample of complete cases, so the RCT was under-powered and unbalanced. Further, the IPE identified that the lack of structure inherent in WfA meant it was not possible to identify fidelity to a programme. This, as well as poor adherence to treatment and control allocations, and concerns about the identification of eligible pupils, mean that the findings of the RCT are presented with low confidence.



References

d'Ydewalle, G., Praet, C., Verfaillie, K., and Van Rensbergen, J. (1991) "Watching subtitled television: Automatic reading behaviour". *Communication Research* Vol. 18(5): 650–666.

Frayer, D., Frederick, W., and Klausmeier, H. (1969) *A schema for testing the level of cognitive mastery*. Working Paper No. 16 Wisconsin Research and Development Center. Madison: University of Wisconsin.

Harvey, S., and Goudvis, A. (2000). *Strategies that work: Teaching comprehension to enhance understanding*. York, ME: Stenhouse.

Jones, K., & Burns, A. (2021) *Unit costs of health and social care 2021*. Canterbury: Personal Social Services Research Unit, University of Kent.

Miller, M., and Veatch, N. (2011) *Literacy in context: Choosing instructional strategies to teach reading in content areas for students grades 5–12*. United States: Pearson.

NASUWT (2022) *England pay scales*. Available at:

<https://www.nasuwt.org.uk/advice/pay-pensions/pay-scales/england-pay-scales.html> (Accessed: 9 September 2022).

Parkhill, F., and Johnson, J. (2009) "An unexpected breakthrough for rapid reading improvement: AVAILL uses movies so students read it, see it and get it". New Zealand: Council for Research.

Quigley, A. (2020) *Closing the reading gap*. Abingdon: Routledge.

Sanders, M., et al. (2020) *What works in education for children who have had social workers?* Technical report. Available at:

https://whatworks-csc.org.uk/wp-content/uploads/WWCSC_what_works_education_children_SWs_Technical_Report_Feb20-1-A.pdf (Accessed: 9 September 2022).

Styles, B., Stevens, E., Bradshaw, S., and Clarkson, R. (2014) *Vocabulary enrichment intervention programme: Evaluation report and executive summary*. Slough: National Foundation for Educational Research.

Walker, W., Worth, J., & Van den Brande, J. (2019) *Teacher workload survey 2019*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/855933/teacher_workload_survey_2019_main_report_amended.pdf (Accessed: 9 September 2022).

Youth in Mind (2016) *Scoring the strengths & difficulties questionnaire for age 4–17 or 18+*. Available at: <https://www.sdqinfo.org/py/sdqinfo/c0.py> (Accessed: 9 September 2022).



Appendix 1: Data tables

Research question 1

Primary analysis

Table 11. Reading skills: complier-averaged causal effect (CACE) analysis table

Variable	Estimate (standard error)
Treatment condition: received WfA	2.72 (5.94)
Time period: post-treatment	1.76 (2.73)
Programme effect: CACE (Received WfA x post-treatment)	5.14 (5.29)
<hr/>	
R ²	0.20
Adj. R ²	0.16
Num. obs.	442
RMSE	22.89

Secondary analysis

Table 12. Reading skills: CACE and intention to treat (ITT) effect comparison table

Variable	CACE estimate (standard error)	ITT estimate (standard error)
Treatment condition: assigned to WfA		1.58 (3.45)
Treatment condition: received WfA	2.72 (5.94)	
Time period: post-treatment	1.76 (2.73)	2.65 (2.05)
Programme effect: ITT (assigned to WfA x post-treatment)		3.00 (3.07)
Programme effect: CACE (received WfA x post-treatment)	5.14 (5.29)	
<hr/>		
R ²	0.20	0.22
Adj. R ²	0.16	0.19
Num. obs.	442	442
RMSE	22.89	22.60



Sensitivity checks

Table 13. Reading skills: CACE sensitivity checks

Variable	CACE estimate Original model (standard error)	CACE estimate 1st check model (standard error)	CACE estimate 2nd check model (standard error)	CACE estimate 3rd check model (standard error)
Treatment condition: received WfA	2.72 (5.94)	3.11 (5.95)	4.47 (6.09)	3.84 (6.16)
Time period: post-treatment	1.76 (2.73)	1.76 (2.73)	1.76 (2.73)	1.76 (2.73)
Programme effect: CACE (Received WfA x post-treatment)	5.14 (5.29)	5.14 (5.29)	5.14 (5.29)	5.14 (5.29)
School fixed effects	Yes	Yes	No	No
Controlling for gender	Yes	No	Yes	No
Controlling for year group	Yes	Yes	Yes	Yes
R ²	0.20	0.19	0.02	0.02
Adj. R ²	0.16	0.16	0.00	0.00
Num. obs.	442	442	442	442
RMSE	22.89	22.91	24.99	25.01

Table 14. Reading skills: ITT sensitivity checks

Variable	ITT estimate Original model (standard error)	ITT estimate 1st check model (standard error)	ITT estimate 2nd check model (standard error)	ITT estimate 3rd check model (standard error)
Time period: post-treatment	2.65 (2.05)	2.65 (2.05)	2.65 (2.05)	2.65 (2.05)
Programme effect: ITT (Assigned to WfA x post-treatment)	3.00 (3.07)	3.00 (3.07)	3.00 (3.07)	3.00 (3.07)
School fixed effects	Yes	Yes	No	No
Controlling for gender	Yes	No	Yes	No
Controlling for year group	Yes	Yes	Yes	Yes



Variable	ITT estimate Original model (standard error)	ITT estimate 1st check model (standard error)	ITT estimate 2nd check model (standard error)	ITT estimate 3rd check model (standard error)
R ²	0.22	0.22	0.06	0.05
Adj. R ²	0.19	0.19	0.04	0.03
Num. obs.	442	442	442	442
RMSE	22.60	22.59	24.57	24.65

Imputation

Table 15. Reading skills: CACE imputation model results

Variable	CACE estimate Model with complete cases (standard error)	CACE estimate Imputation 1 model (standard error)	CACE estimate Imputation 2 model (standard error)
Treatment condition: received WfA	2.72 (5.94)	2.60 (5.17)	2.30 (5.35)
Time period: post-treatment	1.76 (2.73)	2.39 (2.67)	1.30 (2.50)
Programme effect: CACE (received WfA x post-treatment)	5.14 (5.29)	3.85 (4.86)	4.10 (4.58)
School fixed effects	Yes	Yes	Yes
Controlling for gender	Yes	Yes	Yes
Controlling for year group	Yes	Yes	Yes
R ²	0.20	0.20	0.21
Adj. R ²	0.16	0.17	0.18
Num. obs.	442	546	546
RMSE	22.89	21.25	21.73

Table 16. Reading skills: ITT imputation model results

Variable	ITT estimate Model with complete cases (standard error)	ITT estimate Imputation 1 model (standard error)	ITT estimate Imputation 2 model (standard error)
Treatment condition: assigned to WfA	1.58 (3.45)	1.48 (2.69)	0.53 (2.81)
Time period: post-treatment	2.65 (2.05)	4.03 * (1.78)	2.09 (1.62)



Variable	ITT estimate Model with complete cases (standard error)	ITT estimate Imputation 1 model (standard error)	ITT estimate Imputation 2 model (standard error)
Programme effect: ITT (assigned to WfA x post-treatment)	3.00 (3.07)	1.22 (2.59)	2.26 (2.40)
School fixed effects	Yes	Yes	Yes
Controlling for gender	Yes	Yes	Yes
Controlling for year group	Yes	Yes	Yes
R ²	0.22	0.22	0.23
Adj. R ²	0.19	0.19	0.21
Num. obs.	442	566	566
RMSE	22.60	20.80	21.34

Exploratory analysis: dosage indicator

Table 17. Dosage indicator model 1 associations

Variable	NGRT association to dosage indicator model 1
Treatment condition: assigned to WfA	-2.74 (5.54)
Hours of treatment exposure by school	-0.01 (1.88)
Dosage indicator: (assigned to WfA x Hours of exposure)	1.93 (2.41)
Controlling for sex	Yes
Controlling for year group	Yes
R ²	0.10
Adj. R ²	0.02
Num. obs.	97
RMSE	17.94

Table 18. Dosage indicator model 2 associations

Variable	NGRT association to dosage indicator model 2
Treatment condition: assigned to WfA	-0.40 (4.34)



Variable	NGRT association to dosage indicator model
	2
Intensity of treatment exposure by school: high (low intensity = 0)	6.80 (5.26)
Dosage indicator: (assigned to WfA x high-intensity exposure)	-0.08 (8.34)
<hr/>	
Controlling for sex	Yes
Controlling for year group	Yes
<hr/>	
R ²	0.10
Adj. R ²	0.02
Num. obs.	97
RMSE	17.92



What Works *for*
**Children's
Social Care**

info@whatworks-csc.org.uk

 [@whatworksCSC](https://twitter.com/whatworksCSC)

whatworks-csc.org.uk

