HELPING HANDWRITING SHINE

Evaluation Report

November 2020
Updated: September 2022

Gemma Stone, Joana Andrade, Kerry Martin, Ben Styles

The Education Endowment Foundation (EEF) is an independent grant-making charity dedicated to breaking the link between family income and educational achievement, ensuring that children from all backgrounds can fulfil their potential and make the most of their talents.

The EEF aims to raise the attainment of children facing disadvantage by:

- identifying promising educational innovations that address the needs of disadvantaged children in primary and secondary schools in England;
- evaluating these innovations to extend and secure the evidence on what works and can be made to work at scale; and
- encouraging schools, government, charities, and others to apply evidence and adopt innovations found to be effective.

The EEF was established in 2011 by the Sutton Trust as lead charity in partnership with Impetus Trust (now part of Impetus - Private Equity Foundation) and received a founding £125m grant from the Department for Education. Together, the EEF and Sutton Trust are the government-designated What Works Centre for improving education outcomes for school-aged children.

For more information about the EEF or this report please contact:

Jonathan Kay
Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
SW1P 4QP

0207 802 1653

jonathan.kay@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Contents

About the evaluator

The project was independently evaluated by a team from the National Foundation for Educational Research (NFER). The principal investigator for this trial was Dr Ben Styles, Head of NFER's Trials Unit. Gemma Stone, Research Director, led the evaluation team.

Contact details:

National Foundation for Educational Research
The Mere
Upton Park
Slough
Berkshire
SL1 2DQ

p: 01904 567647
e: g.stone@nfer.ac.uk

Acknowledgements

We would like to acknowledge the time and input from No More Marking which provided the platform for the comparative judgment processes, and its expertise and assistance during the project.

Report additions

The analysis carried out for the original report published in November 2020 used primary data collected for the evaluation and no data from the National Pupil Database. Given delays accessing the NPD, analysis that required it has been added in the following places and the report republished:

- Year 2 experiment minimum detectable effect size (Table 17)
- Year 2 experiment FSM6 subgroup analysis (Tables 33 and 34)
- Comparison of FSM6 with school-reported FSM eligibility for the Year 5 experiment (Table 35).

Executive summary

The project

This project aimed to train primary teachers and teaching assistants (TAs) to use occupational therapy approaches, to support children's handwriting, and evaluate whether the approach could improve the overall quality of children's writing. The programme was designed and delivered by a team from the School of Psychology at the University of Leeds and the Bradford Institute of Health Research. In this trial, the programme was implemented as a universal approach in Year 2 (age six to seven) and a targeted approach in Year 5 (age nine to ten). The Year 5 children eligible to receive the intervention were slow and effortful hand writers or those with legibility issues unable to read their own handwriting.

The eight-week intervention was spread across two four-week blocks on either side of the Christmas break, with three 30-minute sessions per week. Each session followed a standard structure consisting of three elements: preparing for handwriting, a warm-up pencil control activity, and an explicit handwriting activity. Sessions also integrated metacognitive approaches which encouraged children to plan and evaluate their work. Training for staff included one full-day session (five to six hours), with follow-on support available based on individual school need. A second half-day of training was added to give support with embedding the approach between the end of the eight-week programme and post-testing at the end of the school year.

The research consisted of two randomised controlled efficacy trials in 103 schools: a Year 5 experiment where treatment and control pupils were within the same school (371 pupils) and a Year 2 experiment where the comparison group were drawn from different schools (3,854 pupils). The process evaluation involved case studies in twelve treatment schools (which used interviews and observations), observations of the training for staff, and analysis of intervention delivery logs. The trial took place in schools between June 2018 and July 2019.

Key Conclusions	
1.	Children in the Helping Handwriting Shine (HHS) schools who were in Year 2 and experienced the universal intervention made no additional progress, on average, in their overall writing ability compared to children in the control group. The range of possible impacts for the universal programme in Year 2 includes small negative effects of two months' less progress and small positive effects of up to one month's progress. This result has a high security rating.
2.	Children in the HHS schools who were in Year 5 and experienced the targeted intervention made the equivalent of two additional months' progress in their overall writing ability compared to children in the control group. The range of possible impacts for the targeted programme in Year 5 includes small negative effects of one month's less progress and moderate positive effects of up to four months' progress. This result has a moderate to high security rating.
3.	Children in the HHS schools, either in Year 2 or Year 5, made, on average, no additional progress in writing composition. This result may have lower security than the overall findings. Exploratory analysis also suggests that it is unlikely the HHS programme increased or decreased children's handwriting speed.
4.	Year 2 children in HHS schools who were eligible for Free School Meals (FSM) made, on average, no additional progress in their overall writing ability compared to similar children in the control group. Year 5 children in HHS schools who were eligible for FSM made, on average, two additional months' progress in their overall writing ability compared to similar children in the control group. These analyses are exploratory and have a lower security than the headline findings given the smaller subsample of children included (Y2 FSM n = 774 and Y5 FSM n = 91)
5.	Adherence to the eight-week programme was, on average, medium to high. There were some potential limitations with the delivery logs, meaning these findings should be viewed with caution. Staff and pupils viewed the programme positively, noticing improvements in children's handwriting during the eight-week intervention. Staff were of the opinion that the programme had improved their ability to teach handwriting and that it was relatively easy to implement.

EEF security rating

The findings from the Year 2 trial have a high security rating and the findings from the Year 5 trial have a moderate to high security rating. The security of the Year 2 trial was slightly compromised by the fact 11% of pupils who started the intervention were not included in the final analysis. The Year 5 trial received a lower rating because problems with the within-school design led to contamination reported by both teachers and developers. This was an efficacy trial, which tested whether the intervention worked under developer-led conditions in a number of schools.

Additional findings

Pupils in Year 2 in the HHS schools made, on average, no additional progress in their overall writing ability (as measured by comparative judgement scores on a Writing Assessment Measure) compared to children in the control group. This result is our best estimate of impact, which has a high security rating. As with any study, there is always some uncertainty around the result: the range of possible impacts for the universal programme in Year 2 includes small negative effects of two months' less progress and small positive effects of up to one month's additional progress. Pupils in Year 5 who received the HHS programme, on average, made the equivalent of two months' additional progress in their overall writing ability compared to children in the control group. This result is our best estimate of impact, which has a moderate to high security rating. The range of possible impacts for this result includes small negative effects of one month's less progress and moderate positive effects of up to four months' progress. Due to the statistical uncertainty around this finding, the evaluation team do not consider it to constitute evidence of promise.

The theory of change (ToC) for this programme hypothesised that the intervention would support pupils to produce fast, accurate, and legible writing. The existing evidence base suggests improvements in handwriting would free up cognitive resources for writing composition, which leads to overall improvements in writing ability. Measures of accuracy and legibility were not undertaken here, but exploratory analysis of handwriting speed and writing composition suggests, on average, these aspects of the ToC did not improve. The planned, more granular analysis of the Handwriting Speed Test was unable to be completed. There may have been an impact on handwriting speed that analysis lacked the sensitivity to detect; however, the collected data suggests handwriting speed did not change in an appreciable way.

Some staff interviews suggested improvements were noticeable in children's writing during the eight-week intervention period, but these were not sustained over the longer term. Delivery logs partially explain this trend as adherence to the eight-week programme—in terms of the number of sessions, their length, and the content delivered—was, on average, medium to high. However, the few logs returned during the four- to six-month 'embedding' period that followed the core intervention showed a tailing off of the use of the programme. This suggests an assessment of handwriting speed immediately after the eight-week intervention may have potentially seen a different result.

Staff perceptions of the intervention were very good. It was described as 'universally beneficial' and easy to administer due to the minimal resources and time required for preparation. Staff interviewed reported positive benefits, including increases in handwriting fluency and speed, accuracy and presentation, and handgrip and control. They reported that pupils had a more positive attitude to handwriting and an appreciation of its significance in their school careers. They also reported that their own ability to teach handwriting had improved, owing mainly to a highly structured, clear methodology.

Cost

The cost of Helping Handwriting Shine for a single-form entry school where the programme is used over three years with a whole class is £27 per child. If the programme is used as a targeted intervention then the cost ranges from £172 to £180 per child, depending on whether a teacher or teaching assistant is trained. Staff are expected to spend an hour and a half per week delivering Helping Handwriting Shine—a total of 12 hours over eight weeks. Additionally, staff are expected to spend five to six hours in training.

Impact[1]

Table 1: Summary of impact on primary outcome(s)

Outcome/ Group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	No. of pupils	p-value	EEF cost rating
Writing Assessment Measure (CJ) – Year 2	−0.02 (−0.13, 0.1)	0	🔒🔒🔒🔒🔓	3,421	0.77	£ £ £ £ £
Writing Assessment Measure (CJ) – Year 5	0.12 (−0.05, 0.29)	2	🔒🔒🔒🔓🔓	313	0.16	£ £ £ £ £

[1] This work was produced using statistical data from ONS. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates.

Introduction

Background evidence

As children learn to communicate via mark-making during their first years in education, many will encounter physical, psychomotor, and sensorimotor barriers that can inhibit the development of handwriting as a skill, which can have wide-ranging consequences on educational attainment (Wallen et al., 2013). There is a solid body of evidence to suggest that there is a percentage of pupils who struggle with fine motor skills and consequently have poor handwriting (Preston et al., 2017). Approximately 10% of all pupils have this type of impairment which 'interferes with their daily living' and is therefore considered a disability in health settings (Diagnostic and Statistical Manual of Mental Disorders, 2013). The prevalence of such problems is higher among pupils who are socio-economically disadvantaged (Liu et al., 2015).

The most prevalent model for supporting these pupils involves referral to occupational therapy (Hoy et al., 2011); the pupil's needs are assessed using standardised diagnostic criteria, which in turn lead to a specific model of support that meets the pupil's requirements. However, waiting lists for referral to Occupational Therapy services are often subject to significant delays, taking up to four years in some areas (Dunford and Richards, 2003). This intervention is based on the premise that schools are well equipped to help children who struggle with fine motor skills and are well placed to take on some of the responsibility of intervention in these scenarios. They also potentially benefit from improving a child's handwriting ability.

The Helping Handwriting Shine programme is designed to address this specific need. It draws on the evidence from systematic reviews of clinical literature (for example, Smits-Engelsman et al., 2013; Preston et al., 2017) that show that for children with clinical motor deficits such as Developmental Coordination Disorder, task-oriented approaches are more effective than process-oriented approaches (that is, the child is taught how to complete a specific task rather than being treated for general 'process' deficit). Other aspects that form part of the intervention are the Cognitive Orientation to daily Occupational Performance (CO-OP) approach, which encourages a pupil to identify directly a task that they want to improve or an aspect of a task that they are finding difficult; to plan how to do the task, reflect, and refine it (Smits-Engelsman et al., 2013; Pless and Carlson, 2000). Occupational therapy techniques, such as positioning and ergonomics, are also part of the intervention.

Age of the student has also been shown to be relevant. In several studies where children aged around six or seven are the target population, they are shown to have a more significant correlation between handwriting automaticity and quality of writing composition than children aged around ten or eleven, comparatively (Medwell et al., 2009, Berninger et al., 1998). As Medwell et al, posit, this 'may indicate that as writers develop, and write more sophisticated texts, there are other issues which account for more of the variance'. As such, core, homogenous, class-based intervention without individualisation at the student level is less likely to have a significant impact on older children whose 'issues' are less likely to be captured and resolved by a one-size-fits-all approach.

The developers of the intervention propose that effective delivery of the intervention will lead first to improvements in legibility, which will then lead to improvements in speed and fluency; additionally, they hypothesise that as writing becomes automatic, it will lead to the freeing up of valuable cognitive resources (McCutchen, 1996) and, therefore, improve the child's capability to undertake cognitively effortful behaviour (McCutchen, 1996; McCarney et al., 2013). It has been shown that the ability to produce handwriting automatically improves handwriting speed (Medwell et al., 2009), and both the speed and the increased automation itself enable the writer to produce higher-quality writing composition (Medwell et al., 2009, Kent et al., 2016). Therefore, the quality of overall writing composition is also hypothesised to improve following participation in the intervention.

The EEF's literacy guidance reports highlight the importance of ensuring fluent handwriting (EEF, 2016b; 2017). However, at present there are no commercially available programmes with secure evidence of effectiveness (Evidence4Impact, 2018). The developer team conducted an implementation pilot study of the intervention in ten primary schools involving 515 children who were aged four to eleven (Shire et al., in press). Feedback and recommendations gained during the study were used to refine the intervention, for example, with regard to session duration, group size, resource availability, and age-differentiation of tasks. The study reported that (1) the children found the tasks enjoyable, (2) the background and booklet instructions were easy to understand, and (3) there was a need for more comprehensive staff training.

Given the location of the developer team, in Leeds, it was felt that the project would be a good fit for the North East Primary Literacy Campaign.

The trial was designed as two separate experiments; a Year 2 (six to seven years old) experiment and a Year 5 (nine to ten years old) experiment. The Year 2 experiment is a two-armed randomised controlled trial, randomised at school level, with all Year 2 children participating in the trial, in order that the intervention could be delivered at class level, enabling better embedding into usual practice. The Year 5 experiment was then randomised at pupil level within the schools randomised to the intervention arm of the Year 2 trial. Year 5 pupils were selected using eligibility guidance provided by the developer team then randomised to create an intervention and control group within each intervention school in the Year 2 trial. By combining the experiments into one set of schools, we were able to maximise the sample size and accommodate the Year 5 pupil targeting where it would be unwieldy to recruit and randomise at school/class level. This design yields considerable efficiencies in delivering the training to all teachers/staff associated with Year 2 pupils and targeted Year 5 pupils within the same school; it also allowed better coverage across both year groups within the case studies and other IPE activity. Furthermore, it achieved considerably more statistical power at Year 5 than a school-randomised design.

Pre and post measurement was carefully considered due to the difficulty of separating an assessment of quality of overall writing composition from an assessment of handwriting. The selected measurement design for the primary outcome was the use of an open writing prompt derived from the Writing Assessment Measure (Dunsmuir et al., 2015) with responses assessed using comparative judgment, which allows a holistic assessment of overall writing quality. Comparative judgment is reliable (Pollitt, 2012), faster, and cheaper than conventional marking and ideal for the large amount of marking that the school-level randomisation generates. For the secondary outcomes, a subsample of scripts from Year 2, and all Year 5 scripts, were transcribed to typescript and then marked a second time using a slightly adjusted version of the published, criterion referencing-based mark scheme for the Writing Assessment Measure (see Appendix M) with the handwriting element removed. This enables an overview of change with any potential handwriting bias removed.

The implementation and process evaluation were designed to address the key points of the theory of change, particularly focusing on the multiple steps identified (that is, first to improve handwriting itself, then that the handwriting becomes automatic and so frees up cognitive space, then that the additional cognitive space allows for better writing composition). Researchers were able to follow the journey of the intervention from start to finish: from observation of the training sessions, through deep dive interviews and case study observations in certain schools, and tracking the delivery and schools' adherence and engagement via a set of monitoring logs; this was bookmarked with a set of interviews with the developers at both commencement and conclusion of the intervention. The activities allowed researchers to gain evidence, at both breadth and depth, of the experiences of schools and pupils, as well as gaining an understanding of any barriers to success in the intervention programme or via the trial design.

Intervention

The Helping Handwriting Shine (HHS) intervention was developed under the supervision of Professor Mark Mon-Williams from the School of Psychology at the University of Leeds and Bradford Institute of Health Research. The delivery core team consisted of a coordinator, one postdoctoral research fellow (PDRF), and one clinical specialist/academic. This team was based at the University of Leeds and at the Bradford Institute of Health Research. The NFER recruited schools to the trial in addition to running the evaluation.

The intervention consists of a set of handwriting practice materials, contained in a programme handbook, that are designed to be delivered over eight weeks in short sessions lasting 30 minutes, three times per week. It can be delivered by one or more trained staff members within a school. Before the commencement of the eight-week delivery period, teachers and support staff in intervention schools were required to attend a training session preparing them for delivery; at the training, they received the programme handbook which sets out all activities plus a suggested timetable matching activities to sessions (though teaching staff could mix and match warm-ups and handwriting activites). Each session consists of three elements:

- preparing for handwriting—in which pupils prepare themselves for the session by addressing their seated position and handgrip;

- a warm-up activity—in which pupils complete a short pen/pencil task that involves practising the types of skills pupils need for writing, such as following a dotted line; and
- a handwriting activity—a longer task where pupils directly practice handwriting skills such as repeating a single letter or a bigram (a pair of consecutive letters).

Activities are of differing levels of difficulty, up to difficulty level three. The session must be based around a 'model, plan, evaluate' structure, a metacognitive concept that requires the teacher to model the activity first, engage the pupil in considering how they may approach the task, then evaluate how the task went once it is complete. There was a suggested model-plan-evaluate table for every activity. 'Gap tasks' are then provided in the handbook consisting of short activities carried out between sessions that reinforce and expand the learning of the previous main session. Extension tasks are also provided for each activity, going beyond difficulty level three. For Year 2, it was anticipated that whole classes were taught together, though possibly grouped within the class. For Year 5, it was recommended that pupils who were randomised to receive the intervention were taught separately in order to eliminate contamination with the control group.

The minimum expectation for schools was that the intervention would be delivered over eight weeks, spread across two four-week blocks: one before Christmas 2018 and one after, avoiding the last week before the Christmas holidays and the first week of the new term in January 2019. In agreement with the developer team, schools were able to amend their delivery to suit them, for example, some schools did not start back on the intervention for two weeks after the holidays. The three thirty-minute sessions were required to be delivered separately, on different days of the week. These requirements were not flexible.

In addition to the formal intervention, additional support activities were introduced by the developer team to cover the period between the end of the formal eight-week intervention period and the testing. These activities were not stipulated in any of the initial documentation of the trial (such as the Memorandum of Understanding or school information sheets), however, they were explained to schools at the full-day training at the commencement of the intervention. They included a half-day training session to help schools transition into normal teaching, and supporting schools in creating an action plan for this transition and the following period until the end of the school year. Communication between the developer team and schools was weekly during this period. Additionally schools were invited to design a gap or extension task as part of a competition; these were shared in newsletters with schools. Eleven newsletters were sent between 1 March and 24 May 2019: the first ten newsletters contained four gap tasks and the last newsletter contained eight (48 in total). In a final additional competition, schools were invited to vote on the best submissions, with the two schools with the highest votes winning a prize.

Some variability was predicted in the extent to which schools would engage with the intervention during this period. The IPE was designed to monitor this period in addition to the formal intervention period.

Some adaptation to the programme was deemed acceptable by the developers but core features were emphasised within the training and handbook. Five underpinning principles, which outlined what could not be changed, were made clear in both the training event and the handbook:

1. frequency of practice;
2. structure;
3. variability of practice;
4. metacognition; and
5. engaging with support.

In addition, five adaptable aspects were outlined in the training and the handbook:

1. handwriting scheme and styles;
2. levels;
3. extension tasks;

4. gap tasks; and
5. rewards.

The evaluation and developer teams both identified the potential for contamination within Year 5 classes: despite the requirement that Year 5 sessions should be delivered outside of class, there remained a risk that these pupils would receive their support within the same classroom. There was also a wider contamination concern, even when intervention pupils were taught separately, that if control pupils were taught by a teacher who had received the intervention training, that teacher may (consciously or subconsciously) pass on some of the techniques to control pupils. Mitigations put in place for this were:

- the intervention was delivered by a member of staff other than the class teacher;
- pupil names were printed onto their intervention learning materials;
- schools were not given spare copies of the materials;
- NFER provided guidance for the training team on how to advise schools about avoidance of contamination during the training sessions; and
- NFER produced a 'dos and don'ts' poster for schools to display in staff areas (see Appendix J2).

Recruitment and delivery

The target regions for the programme were Leeds, Bradford, Wakefield, Barnsley, Doncaster, Rotherham, Sheffield, and the North East (NE). All Year 2 pupils within the target regions who attended primary schools with a Year 2 and Year 5 were eligible to receive the intervention.

Schools randomised to the intervention arm were required to select between five and fourteen Year 5 children, depending on the number of Year 5 pupils. The number of children that a school could put forward is defined in Table 2.

Table 2: Guidance on number of Year 5 pupils per school

Number of pupils in Year 5	Number of pupils to select
0–29	5
30–59	8
60+	14

This enabled a roughly proportional sample to school size and also enabled the minimum sample size to be met (370) with a reasonable buffer. Schools were given a number of pupils to nominate against a set of criteria (specified below); we were keen to make sure schools understood that if they believed they had more, or indeed fewer, pupils meeting the criteria than the number we specified, they should alert us, and provide their data for the trial accordingly. This happened in just two schools. One school had chosen 13 instead of 14 and another chose 15 instead of 14.

The developer team defined two groups of Year 5s eligible to receive the intervention: slow and effortful hand writers and those able to write faster but unable to read their own handwriting. Children whose handwriting is messy but fast/non-effortful and legible were not eligible. Schools were required to rank and select their pupils according to how much additional support they need, applying these criteria. The criteria were supplied for schools by the developer and guidelines were developed for schools regarding the selection of pupils. Schools were asked to contact the developer directly if they had queries about the selection process. None did so, but some did express in training events that they had found this difficult. Once selection had taken place, the selected Year 5 pupils took the baseline assessment and were randomised into control and intervention groups within each school.

The trial recruited 103 schools in total. One school withdrew from the Year 5 experiment only but fully participated in the Year 2 experiment. Seven schools withdrew from the Year 2 experiment intervention but still completed the baseline and endpoint tests. No schools withdrew from the control group in the Year 2 experiment.

Training took place in October and November 2018 and was delivered by a team comprised of one post-doctoral research fellow (PDRF) and one research fellow, employed by the University of Leeds (with expertise in psychology and occupational therapy). This team was supported by two external experts in continuing professional development and specialists in health training from the Bradford Institute of Health Research (an NHS research organisation).

The training events were held in each of the following geographical areas: Leeds, Sheffield, Darlington (North East, south) and Newcastle (North East, north). Two full-day (five to six hours) training sessions ran in each location, with each participating school required to attend one full-day session. A minimum of two members of staff were expected to attend with a maximum of five members in total. Staff could include not only teachers but also teaching assistants (TAs), SENCOs, senior leaders, or occupational therapists. The schools were required to identify staff who could deliver the intervention across Year 2 and in Year 5; it was recommended that staff delivering to Year 5 were not teachers but rather other staff (as listed above) that were able to take the targeted pupils out of class to receive the intervention while normal teaching continued in the classroom with the teacher. Cascading was not a recognised feature of the intervention and therefore schools were encouraged to send to the training all those delivering the intervention.

Following the training, some support was anticipated in the form of follow-up site visits or phone support for 'troubleshooting' (with a mixed approach, depending on need, during the course of the intervention and beyond). The developer team anticipated 'weekly check-ins' with schools.

Issues arising during the project

Two operational issues arose during the project: firstly, NFER did not collect free school meals (FSM) data from schools for Year 2 as had initially been planned—this information was derived from the NPD instead. Secondly, the initial timeline did not allow enough time for the marking to be done, so this was added in later and a delayed timetable agreed for analysis and reporting. In order to anticipate and ratify any issues with the comparative judgment process and use of the associated platform, we ran a pilot of the assessment process in order to ensure we could run it at scale. This was helpful in ironing out issues that arose, such as placement of QR codes on scripts.

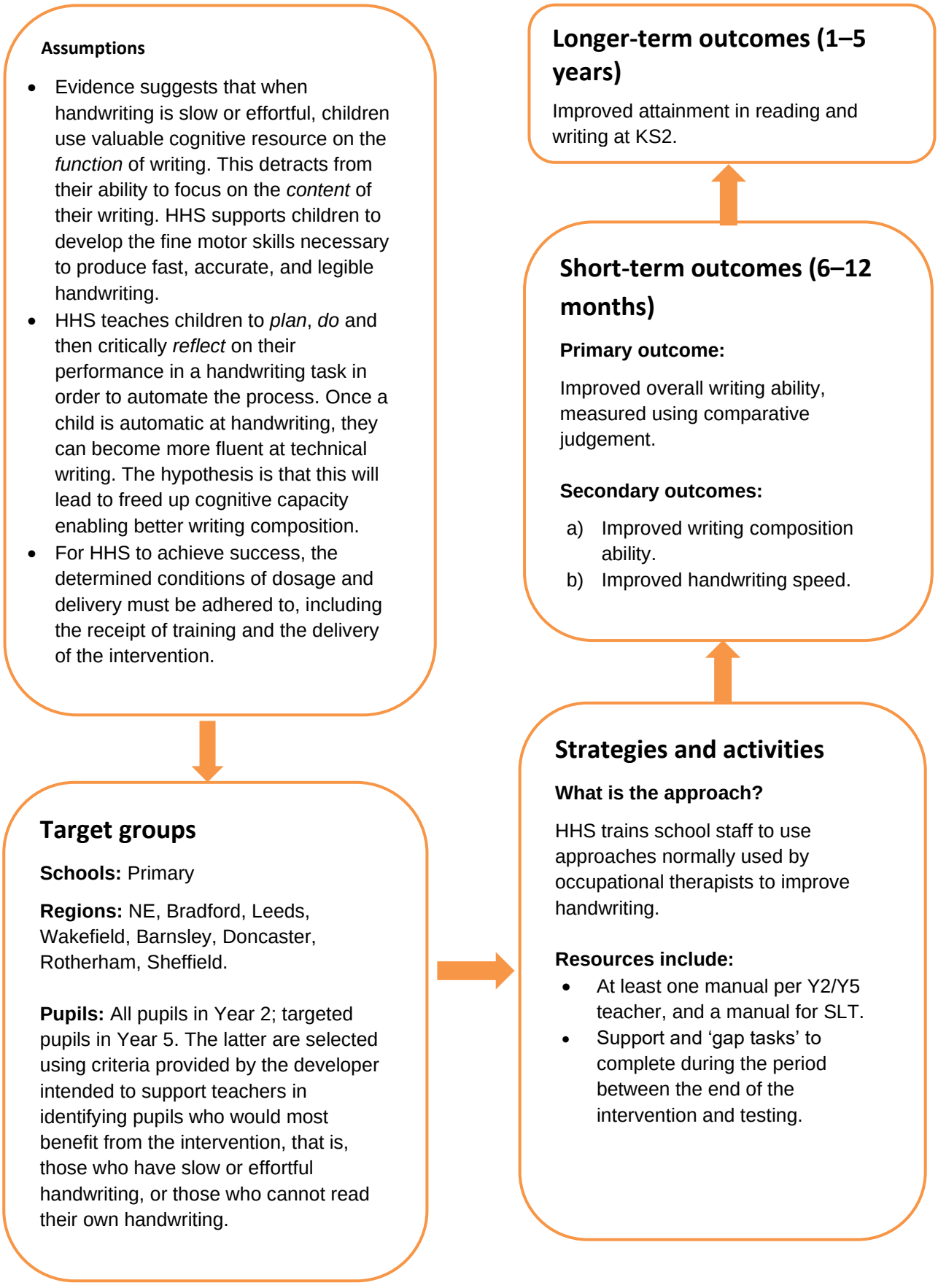
For schools, an issue that became clear early on was that they had not understood, despite being clearly advised, that only half the nominated Year 5 pupils would receive the intervention; they found it difficult to accept that the selection of pupils to receive it had to be random. Feedback that schools were unhappy with this came very quickly from schools when we released the results of the Year 5 randomisation. In addition, because the numbers of Year 5 intervention pupils in some schools was very small, these schools found it harder to justify the additional burden of finding separate teaching rooms and releasing staff for the sessions and for the training.

Another issue for schools was delivery time. Half the intervention was planned to be delivered before the Christmas break and half afterwards, but this was problematic for some schools: they reported finding it difficult to fit in the sessions within the few weeks before the holidays due to additional school commitments such as shows, trips, and so forth. Similarly, returning in January, there was a delay for some schools in restarting the intervention due to other commitments. On receiving this information, we encouraged schools to restart the intervention as quickly as possible but also gave them the flexibility to deliver whatever remained of the intervention, as opposed to only delivering what they could within the specified time period. This meant that all schools completed the intervention, but had the knock-on impact that some schools finished delivering the sessions early in February and some not until late March. Therefore, the amount of time each school had left to 'embed' the techniques and approaches from the programme was different.

For the evaluator, we began to understand that the design of the Year 5 experiment would make it difficult to embed the intervention in the period between the end of the formal intervention and the testing due to the risk of—and strategies to minimise—contamination of the control group (see Appendix J2). Similarly, the definition of the embedding period—its purpose, requirements for schools, and monitoring of the activity—was not defined clearly enough at commencement of the trial. Clear monitoring would have mitigated this to some extent, giving information on what had happened to allow a greater understanding of how this related to outcomes. Unfortunately, return on the monitoring logs intended to capture activity and engagement during this period was very low, though the developer team achieved good engagement from schools during this period.

Theory of Change

Figure 1: Helping Handwriting Shine theory of change



Evaluation objectives

Impact evaluation research questions

Primary questions

RQ1a: What is the impact of the Helping Handwriting Shine intervention on the comparative judgement measurement scale for writing of children aged six to seven years old?

RQ1b: What is the impact of the Helping Handwriting Shine intervention on the comparative judgement measurement scale for writing of targeted children aged nine to ten years old?

Secondary questions

RQ2a: What is the impact of the Helping Handwriting Shine intervention on handwriting speed of children aged six to seven years old?

RQ2b: What is the impact of the Helping Handwriting Shine intervention on handwriting speed of children aged nine to ten years old?

RQ3a: What is the impact of the Helping Handwriting Shine intervention on writing composition of children aged six to seven years old?

RQ3b: What is the impact of the Helping Handwriting Shine intervention on writing composition of children aged nine to ten years old?

RQ4: Are effects on writing ability (as indexed by RQs above) different for pupils eligible for FSM? If so, how?

RQ5: Is there an interaction between fidelity and attainment for treatment schools?[2]

Implementation and process evaluation research questions

RQ1: Is fidelity to the intervention maintained?

RQ2: How much does dosage differ across the sample?

RQ3: To what extent do participants engage with the intervention?

RQ4: What level and type of support does the developer provide to intervention schools?

RQ5: What does 'business as usual' (BaU) consist of?

The project protocol is available at:
https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/Handwriting_evaluation_protocol_UPDATED_121118.pdf

The project's statistical analysis plan (SAP) is available at:

https://educationendowmentfoundation.org.uk/public/files/Projects/HelpHandwritingShine_SAP_vFinal_10.05.2019.pdf

[2] Please note the wording of this question differs slightly from the published SAP, but is reflected in the study protocol.

Ethics and trial registration

The trial was designed, conducted, and reported to CONSORT standards (http://www.consort-statement.org/consort.statement/) and registered on http://www.controlled-trials.com/. The evaluation was conducted in accordance with NFER's Code of Practice, available at: http://www.nfer.ac.uk/nfer/about-nfer/code-of-practice/nfercop.pdf. NFER, the University of Leeds, the EEF, and No More Marking worked closely together to ensure each organisations' policies could be applied in practice.

Ethical agreement for participation within the trials was provided by the headteacher of the school via a signed Memorandum of Understanding (see Appendix I). Parents were provided with full details about the intervention and were given the opportunity to withdraw their child from data processing if they had objections to this (see Appendix L). Participant opt-in consent was sought for participants (teachers, parental consent for children) in the interviews/case studies that formed the IPE. Parental consent was gathered prior to the interviews via letter.

The ISRCTN trial registration number is ISRCTN13315075.

Data protection

NFER collected personal data to enable the evaluation of the Helping Handwriting Shine (HHS) intervention using a randomised controlled trial. Personal data was collected by NFER directly from schools and through matching to the National Pupil Database. This was augmented by assessment data provided both by the schools and through visits by NFER test administrators. The University of Leeds collected registers at training events that captured names and places of work.

Personal data for the trial included data about teachers and pupils from participating schools as described below:

- NFER collected data (name, job title, and contact details) about a nominated named teacher and two staff members within a participating school so that we could liaise with them about the evaluation.
- NFER collected personal data about pupils. This included pupil name, date of birth, unique pupil number (UPN), class name, and school name for all pupils at Year 2 and for nominated pupils at Year 5. For nominated Year 5 pupils, free school meals status was also collected.
- NFER accessed pupil background data held on the Department for Education National Pupil Database (NPD). The NPD data that we requested is FSM eligibility and gender.

NFER matched all of the above pupil data to pupil assessment data. This assessment data included pupil results from (a) two baseline tasks sat in the summer of 2018—the Writing Assessment Measure and the Handwriting Speed Test and (b) from the same two tests sat at the end of the trial in the summer of 2019. NFER shared the test results with schools in October 2019. The above datasets enabled NFER to undertake primary and secondary outcomes analyses for the trial. NFER will share all of the above pupil data (name, date of birth, UPN matched to the NPD data described above, and assessment results) with the EEF's data archive partner, the Fischer Family Trust. Anonymised data will also be stored in the UK Data Archive.

Personal data was stored on the No More Marking comparative judgment platform in order to use the platform for marking. As such, senior members of staff at No More Marking had access to this data in their management of the platform, and acted as data processor. Data included the pupil's name, date of birth, school, and UPN. Markers from the company, and from NFER, did not have access to the personal data; test papers were presented anonymously. Pupil data has been treated with the strictest confidence. Neither we, nor any of the named parties, have used pupil names or the name of any school in any report arising from the research.

NFER, the University of Leeds, and No More Marking will delete any personal data before three years from the completion of the project. (Note that retention of personal data is subject to agreement by the NPD team at the DfE.) NFER will send all the data it has collected to be archived, via the Secure Research Service, within three months of project completion.

The archived data will be available in a de-identified form with restricted access for research purposes only via the Secure Research Service. NFER handles personal data in accordance with the rights given to individuals under data protection legislation. Individual rights are respected. No personal data is stored or transferred outside of the EEA.

In setting out the roles and responsibilities for this trial, the four parties (NFER, the University of Leeds, No More Marking, and the EEF) have signed a Data Sharing Agreement. This includes a description of the nature of the data being collected and how it will be shared, stored, protected, and reported by each party. In addition, NFER provided a memorandum of understanding to schools explaining the nature of the data being requested of schools, families, and children, how it will be collected, and how it will be passed to, and shared, with NFER.

Legal basis

The legal basis for processing the personal data accessed and generated by the trial is covered by GDPR Article 6 (1) (f) which states that:

> 'processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party except where such interest are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of the personal data'.

NFER carried out a legitimate interest assessment which demonstrates that the evaluation fulfils one of NFER's core purposes (undertaking research, evaluation, and information activities) and is therefore in our legitimate interest, that processing personal information is necessary for the administration of the randomised controlled trial. We have considered and balanced any potential impact on the data subjects' rights and find that our activities will not do the data subjects any unwarranted harm.

Table 3: Roles with respect to data management

Organisation	Role
National Foundation for Educational Research (NFER)	Joint data controller
University of Leeds	Joint data controller
No More Marking	Data processor
Education Endowment Foundation (EEF)	Data controller (for archive stage—once the data has been submitted to the archive, after the trial has been completed)

For further information, please see the Privacy Notice for the Evaluation of Helping Handwriting Shine, available at **https://www.nfer.ac.uk/pdf/EEFH_Privacy_Statement.pdf** and at Appendix G.

Project team

Table 4: Project team

Name	Institute	Roles and responsibilities
Dr Ben Styles (BS)	NFER	Trial Director, responsible for leading the NFER team and project delivery.
Gemma Stone (GS)	NFER	Trial manager, responsible for overseeing the day-to-day running of the trial and for managing the process evaluation activities and analysis.
Kerry Martin (KM)	NFER	Process evaluation researcher, responsible for carrying out process evaluation activities and analysis.
Kathryn Hurd (KH)	NFER	Test and Schools administration lead, responsible for overseeing recruitment, school contact, and testing.
Dr Frances Brill (FB)	NFER	Assessment advisor, responsible for guiding the team on selection and marking of appropriate assessments.
Dr Joana Andrade (JA)	NFER	Statistician, responsible for statistical analysis.
Dr Chris Wheadon (CW)	No More Marking	Supplied the comparative judgement platform and advised on interpreting and modelling the resulting data.
Prof. Mark Mon-Williams (MMW)	University of Leeds	Lead developer, responsible for delivery of the intervention.
Jo Atkinson (JA)	University of Leeds	Research Fellow and Occupational Therapist, responsible for contributing to the design of the intervention training, delivering to school staff, and supporting schools during the intervention.
Dr Emily Williams (EW)	University of Leeds	Post-Doctoral Research Fellow, responsible for contributing to the design of the intervention training, delivering to school staff, and supporting schools during the intervention.
Dr Katy Shire (KS)	Bradford Institute for Health Research	Developer steering group member, responsible for assisting with intervention development and delivery.
Dr Liam Hill (LH)	University of Leeds	Developer steering group member, responsible for assisting with intervention development and delivery.
Dr Amanda Waterson (AW)	University of Leeds	Developer steering group member, responsible for assisting with intervention development and delivery.
Dr Nick Preston (NP)	University of Leeds	Developer steering group member and physiotherapist, responsible for assisting with intervention development and delivery.
Charlotte Clowes (CC)	The Aspire Educational Trust	Developer steering group member, responsible for consulting on intervention training.
Prof. David Sugden (DS) [Deceased]	University of Leeds	Developer steering group member, responsible for creating and developing the intervention and consulting on intervention training.
John Pickavance (JP	University of Leeds	Developer, steering group member, contributed to design of materials and delivery of teacher training sessions.

Methods

Trial design

Due to the necessity for different modes of delivery to different child age brackets, the HHS trial was designed to include a Year 2 experiment (targeted at six- to seven-year-old children) and a Year 5 experiment (targeted at nine- to ten-year-old children). The Year 2 experiment is a school-randomised efficacy trial that evaluated the impact of class-based intervention on the writing ability of younger children. To reflect the nature of the class-based mode of delivery, this trial involved the whole of Year 2 in each school.

The Year 5 experiment is a multi-site efficacy trial that evaluated the impact of the HHS intervention through small group settings on the writing ability of older children who struggle to handwrite. Children who met the eligibility criteria specified in the Study Design section were selected in each of the intervention group schools of the Year 2 experiment to take part in this trial. The eligibility criteria for the Year 5 trial include two distinct (but not mutually exclusive) groups: pupils who have slow and effortful handwriting and pupils who cannot read their own handwriting. We also explored whether there was a differential impact of the HHS intervention on each of these two groups.

One of the objectives of the HHS trial was also to evaluate whether a higher or lower degree of fidelity to the implementation protocol had an impact on the overall results of the HHS intervention. For this purpose, we also considered whether there was an association between fidelity and the primary outcome for both experiments. Finally, we also considered the differential effect, if any, of the intervention on FSM-eligible pupils.

The control arm condition for each trial was ‘business as usual’. A £500 incentive was available to control schools that completed baseline and endpoint testing, considered by the developer to be approximately equal to 25% of the cost of the intervention.

The primary outcome for both trials was the Writing Assessment Measure (Dunsmuir et al., 2015, pp. 1–18) assessed using comparative judgment. Secondary outcomes were the Handwriting Speed Test (Wallen et al., 1996) and the Writing Assessment Measure assessed using the associated, criterion-referenced mark scheme.

Table 5: Trial design

Trial design, including number of arms		Two randomised controlled trials, each with two arms
Unit of randomisation		Year 2 experiment: school Year 5 experiment: pupil
Stratification variable (s) (if applicable)		Year 2 experiment: region (training hub) Year 5 experiment: school and FSM eligibility
Primary outcome	Variable	Writing ability
	Measure (instrument, scale, source)	Writing Assessment Measure (WAM_CJ) (comparative judgement true scores)
Secondary outcome(s)	Variable(s)	Writing composition Handwriting speed
	Measure(s) (instrument, scale, source)	Writing Assessment Measure (WAM_CR) (criterion reference scores excluding the handwriting element) Handwriting Speed Test (HST_FS) (aggregated to full sentence count)
Baseline for primary outcome	Variable	Writing ability
	Measure (instrument, scale, source)	Writing Assessment Measure (WAM_CJ) (comparative judgement true scores)
Baseline for secondary outcome(s)	Variable	Writing ability Handwriting speed
	Measure (instrument, scale, source)	Writing Assessment Measure (WAM_CJ) (comparative judgement true scores) Handwriting Speed Test (HST_FS) (aggregated to full sentence count)

Participant selection

Recruitment to the trial was conducted by NFER between February and May 2018. Eligible schools were those within the identified regions—Leeds, Bradford, Wakefield, Sheffield, Doncaster, Rotherham, Barnsley, and the North East local authorities—with at least one Year 2 class and one Year 5 class in the academic year 2018/2019. Schools could come from the maintained sector but not from the private sector. The EEF had received Expressions of Interest from schools in the North East with regards to the North East Literacy Programme, and, as such, this list of schools was automatically added to the sample providing they met the eligibility criteria.

NFER contacted the local authority (LA) first to enquire whether they would recommend that certain schools should be excluded from the trial because of particular circumstances. Following confirmation from the LA, NFER wrote to all schools in the sample, sending a School Information Sheet and reply form including the Memorandum of Understanding (MoU; see Appendices I and J). Schools then returned the reply form and MoU to NFER, including the details for a nominated contact within the school and the number of pupils and classes in the targeted year groups (2 and 5).

NFER then requested Year 2 pupil data from all schools eligible to enter the trial, which the school uploaded via a secure portal. Schools that had provided data were then put forward to baseline testing followed by randomisation. Schools were only randomised if they had completed baseline tests with their Year 2 cohort. NFER also provided schools with a letter for parents of Year 2 children, informing them that the school was participating in the research and giving them the opportunity to withdraw from sharing the child's data (see Appendix L).

NFER notified schools of the outcome of randomisation in July 2018; if allocated to the intervention group, schools were asked to provide the details for between five and fourteen Year 5 pupils, depending on the number of pupils in the year group, who met the criteria for receiving the intervention, to be uploaded via secure portal. Year 5 pupils were eligible for the targeted intervention if they were deemed to meet the criteria set by the developer (see background section above). NFER required schools to then provide the parent letter to Year 5 children's parents to let them know the school was participating in the research, giving the parent the opportunity to opt out of the trial, including all data sharing. Following receipt of Year 5 pupil details by NFER, Year 5 pupils undertook baseline testing in schools in September 2018 consisting of the same tasks as for Year 2. Those who took the tests were randomised within the school, with half allocated to the intervention group and half to control. This was notified to schools using the online platform, as per all communication with schools during the trial.

Participation in analysis

Baseline for the primary outcome analysis consisted of all the pupils who attempted the writing assessment and were randomised. Pupils who did not attempt the writing task were later excluded from the analysis. This criterion was applied in both experiments. Baseline for the first secondary outcome analysis consisted, for the Year 2 experiment, of all the pupils randomly selected to be part of the secondary sample and who attempted the baseline writing assessment. As for the primary outcome analysis, only the pupils who attempted the writing task were included in the analysis. An 'attempt' was defined as at least one legible word written on the paper.

The baseline and analysis inclusion criteria for the first secondary outcome analysis for the Year 5 experiment were identical to the criteria considered for the primary outcome analysis. Baseline for second secondary outcome analysis consisted, for the Year 2 experiment, of all the pupils randomly selected to be part of the secondary sample and who attempted the handwriting speed test. Pupils that did not attempt the baseline handwriting speed test were excluded from the analysis. The same exclusion criteria were also applied for the Year 5 experiment: only pupils who attempted the handwriting speed test were included in the analyses.

Exclusion criteria identical to the ones adopted at baseline were also implemented at follow-up: only pupils who attempted the writing and handwriting speed test tasks were included in the analysis. Missing data analysis was carried out as per the SAP.

Outcome measures

All baseline measures were administered by class teachers while all endpoint measures were administered by NFER test administrators. All markers/judges were blinded to condition at pre- and post-test. Teachers, at baseline, and test administrators, at endpoint, were given the same information and instructions for administering the assessment. In both instances they read from a script to ensure that the information presented was identical. Teachers were present during endpoint testing to ensure pupils were comfortable and able to ask questions.

Primary outcome measure

The primary outcome measure for both experiments was the Writing Assessment Measure (WAM) assessed using comparative judgement to measure the 'best writing'; this instrument and assessment methodology was selected to give an overall measurement of the writing construct with high reliability. The assessment was 'judged' (marked using comparative judgment) by external, blind judges derived from a pool of current or former teachers recommended to NFER by No More Marking (NMM). All judges were familiar and confident with both the platform and the judgment process before being recommended to NFER.

In the assessment, pupils were asked to write in response to a prompt. A variety of prompts were available to select from within the WAM materials; the chosen prompt was thought to best reflect the age and ability of the cohorts targeted in the trial:

> *Imagine that you could go anywhere you wanted to on a school trip with your class and your teacher. You could go anywhere at all. Write about where you would go and what you would do.*

Children were given 20 minutes to respond to the prompt, under test conditions. Once received by NFER, scripts were scanned and uploaded to the NMM platform ready to go through the judging process.[3]

Comparative judgment produces a rank score against a set of scripts without reference to any pre-established criteria or norms. The NMM programme randomly selects pairs of scripts from within each 'task' (either Year 2 baseline, Year 2 follow-up, Year 5 baseline, or Year 5 follow-up). Judges are presented with each pair and then asked to choose which of the pair is the better piece of writing (see below for precise instruction). They then click on their choice. The programme continues to present random pairs until a certain number of decisions per script has been reached (this number is selected based on number of judges and scripts and input to the platform before judging begins). Judges were briefed via email prior to beginning judging; their briefing did not include details on the ages or ability of children, nor the purpose of the testing. They were briefed only to select the best piece of writing. When viewing scripts within the system, they could not see any identifying information; only an NFER student identifier number was visible alongside the writing itself, which they could not connect to pupils' age or ability. However, they were seeing pupils' handwriting, which may have influenced decisions on quality.

The NMM platform uses the Bradley-Terry model (Hunter, 2014, pp. 384–406) to produce 'true scores'. True scores measure a latent ability,[4] in this case writing ability, and are computed from the wins and losses of a script (that is, the times it has been selected as 'best', or not) against other scripts.[5] To produce results with a high level of reliability, ten judgements are made per script (known from previous work by No More Marking to produce a very reliable measure; see Pollitt 2012). True scores are measured in a scale that is linear, robust to missing data, has estimates of precision, detects misfit, and the parameters of the objects being measured can be separated from the measurement instrument being used. Although there is no underlying assumption of normality, since true scores measure a latent writing ability, they are generally normally distributed and are standardised to have a mean of zero and a standard deviation of two.

Judging of all four tasks took place over six weeks during July and August of 2019 by ten judges who each judged across all four tasks. Judges were briefed via email one week before beginning the task and were told that: 'The purpose of the test is to assess overall writing ability. We will not be issuing any specific guidance on what this means, or any particular aspect to consider; please give a holistic judgment, bearing in mind that we are testing for overall writing ability.' All judges were monitored during the judging period using the tracking features within the NMM platform. This included 'percentage left click', which monitors each judge to track what percentage of their selections are of the script presented on the left of the screen to ensure that arbitrary judgements are not being made; no judge was outside the threshold of 40–60% left clicks, which would indicate bias to one side. Additionally judge 'infit' was monitored: this is a measure of consistency between judges.[6] Infit at or below 1.2 is predefined as the acceptable threshold, and no judge in any of the four tasks exceeded this (Table 6).

Table 6: Infit of judges

Task	Maximum infit
Year 2 baseline	0.86
Year 5 baseline	1.02
Year 2 endpoint	0.76
Year 5 endpoint	0.83

Secondary outcome measure 1

Secondary outcome 1 utilised the outcomes of the WAM prompt for the primary outcome but instead of using comparative judgment to assess, the scripts were marked using the published WAM mark scheme, a levels-based

[3] https://www.nomoremarking.com/
[4] See Hunter, 2014, pp. 384–406.
[5] The computation of true scores also takes into account the scores of the scripts that the script was judged against (Hunter, 2014, pp. 384–406).
[6] See the No More Marking blogpost on Judge Infit for a full description of its calculation: https://blog.nomoremarking.com/judge-infit-27aec5ede2d

'traditional' criterion referenced mark scheme (see Appendix M). The mark scheme was then adjusted (see Appendix N) by:

- removing the handwriting criterion;
- inserting a '0' mark band for each category so that the markers could indicate where their judgment was that a given script did not meet the criterion for '1' mark (this is obviously implied in the published WAM but not made explicit); and
- omitting—in the category 'Sentence Structure and Grammar'—the bracketed examples given in two instances as it was felt these were not actually examples of the grammatical features cited in the rubric.

All Year 5 scripts, and a randomly selected sample of five scripts per school from the Year 2 primary outcome (sampled by a statistician), were transcribed to remove any bias that handwriting may introduce on how the content and composition were scored. The transcription was completed by trained NFER staff with 10% of the transcription tasks quality assured. The transcribed scripts were then marked using the WAM mark scheme by subject-specialist markers from the NFER marking pool. The markers were trained and standardised in house before completing any marking and supervised by expert marker managers. As the markers only saw transcriptions, they did not see any details indicating allocation of the pupil or school, nor their age, gender, and so forth, meaning they were entirely blinded.

Secondary outcome measure 2

The second secondary outcome measure was, for both experiments, the Handwriting Speed Test (HST; Wallen et al., 1996). This measure aligns to the first step in the theory of change—that the intervention will improve a child's handwriting speed and fluency. It was administered immediately after the primary outcome assessment as part of the same test sitting. In this test, a short phrase,[7] specifically designed to include the key handwriting shapes and all letters of the English alphabet, was presented to pupils (see Appendices Q1 and Q2). Pupils were instructed to copy the phrase out as many times as they could within a three minute time period. Test administrators provided instructions taken from the HST manual (page 17), which states that pupils will be measured on how 'quickly and neatly' they can write; children are also advised: 'Remember it is not a race, just use your normal writing. Write as quickly but neatly as you can.' The test was marked by NFER markers using the scoring criteria provided in the HST manual. Those scripts that were transcribed for the secondary outcome had their HST marked, that is, five per school for the Year 2 trial and all Year 5 scripts. The remaining Year 2 scripts did not have their HST marked for reasons of cost. As this is a three-minute activity and it is not imperative that pupils receive feedback, it was not deemed cost effective to mark all 3,700 Year 2 scripts. The original version of the test is designed to produce raw scores that can be cross-referenced with normative tables; however, in this context, we compared pre and post scores without reference to norms.

Table 7: Outcome measures and tasks

Research methods	Measure	Task	Sample	Marking method	Transcribed
Year 2 baseline	Primary outcome	WAM	All	CJ	
	Secondary outcome 2	HST	5 per school	Mark scheme	
Year 5 baseline	Primary outcome	WAM	All	CJ	
	Secondary outcome 2	HST	All	Mark scheme	
Year 2 endpoint	Primary outcome	WAM	All	CJ	
	Secondary outcome 1	WAM	5 per school	Mark scheme	Yes

[7] The quick brown fox jumps over the lazy dog.

Year 5 endpoint	Secondary outcome 2	HST	5 per school	Mark scheme	
	Primary outcome	WAM	All	CJ	
	Secondary outcome 1	WAM	All	Mark scheme	Yes
	Secondary outcome 2	HST	All	Mark scheme	

The raw scores generated by the Handwriting Speed Test (HST_RS) corresponded to the number of legible letters pupils are able to write during a three minute window when writing at their usual writing speed (Wallen et al., 1996). 'Legible' is closely defined in the guidance provided in the HST test materials and booklet, and those marking were trained and standardised according to this guidance.

Once the trial data became available the project statistician undertook a quick exploratory data analysis and concluded that the baseline and follow up HST_RS data, for both Year 2 and Year 5, displayed multimodal distributions. In both the Year 2 and Year 5 experiment, the project statistician observed modal values that correspond to multiples of 35, the number of letters in the sentence pupils were asked to write as many times as they could in three minutes. Adjacent values to multiples of 35 also displayed high frequency values when compared to non-adjacent values, as can be seen in Figure 2 to Figure 5, below.

Figure 2: Frequency table (ten most frequent values) and histograms for the Year 2 experiment baseline handwriting speed data (HST_RS)

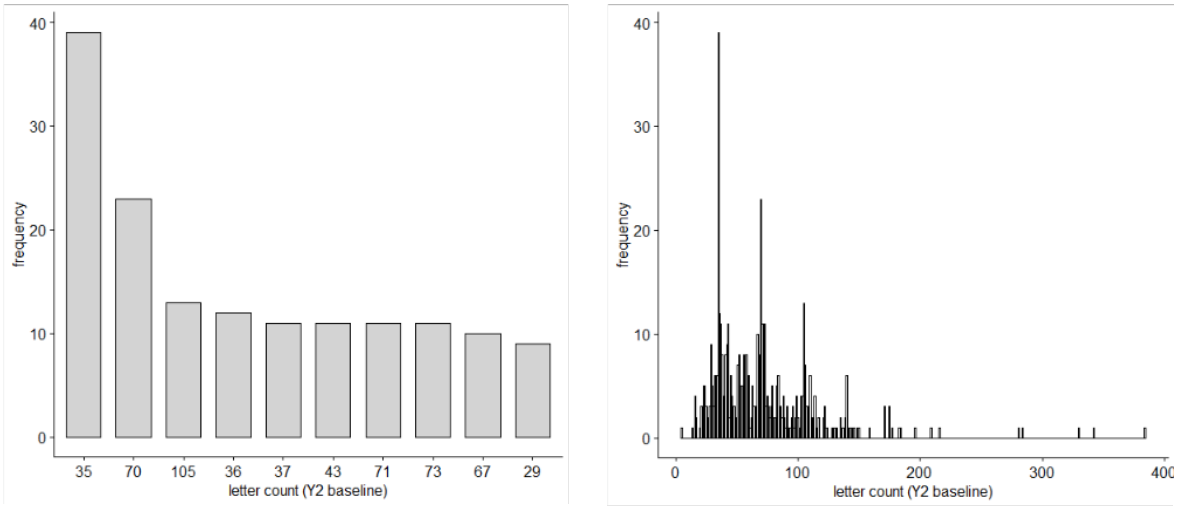


Figure 3: Frequency table (ten most frequent values) and histograms for the Year 2 experiment follow-up handwriting speed data (HST_RS)

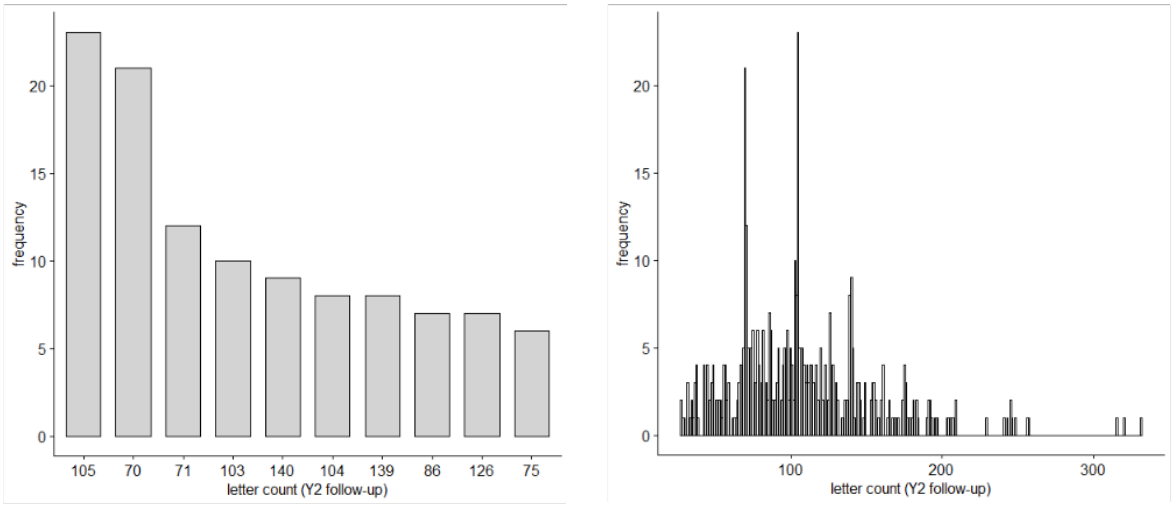


Figure 4: Frequency table (ten most frequent values) and histograms for the Year 5 experiment baseline handwriting speed data (HST_RS)

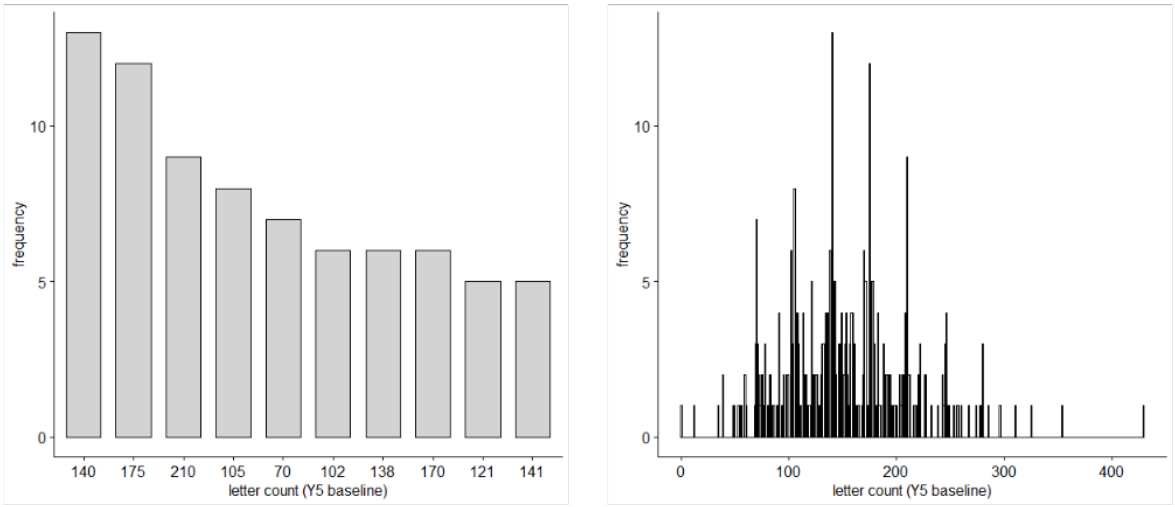
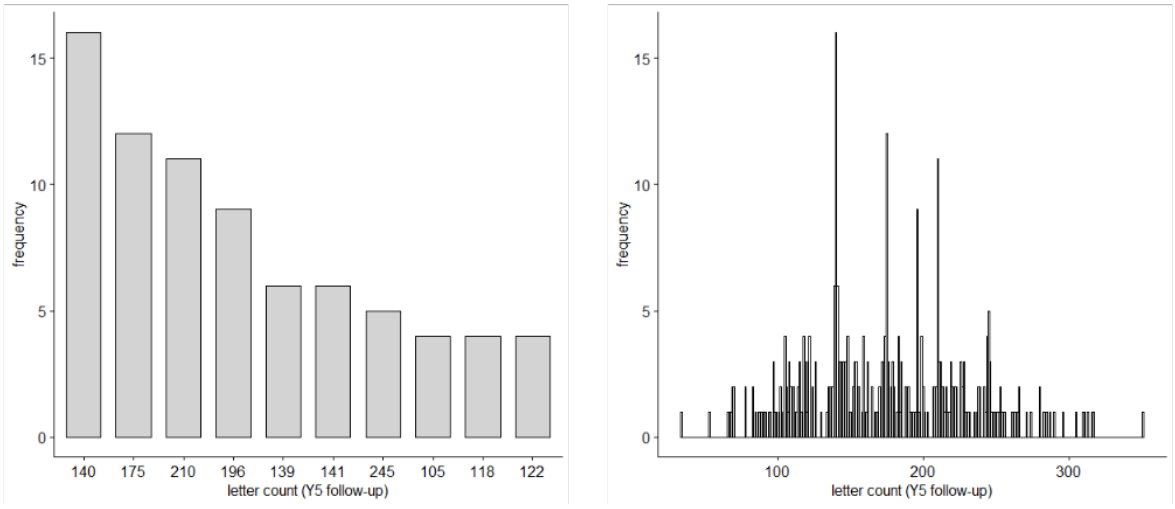


Figure 5: Frequency table (ten most frequent values) and histograms for the Year 5 experiment follow-up handwriting speed data (HST_RS)



The distribution and characteristics of the HST_RS data suggest that, in a substantial number of cases, the instructions for completing the handwriting speed task were understood, either by the administrators of the assessment at baseline or by pupils, to mean that pupils should stop writing at the end of a complete sentence instead of moving on to a new line and writing as many letters as they could within the time limit. Or, that pupils were allowed to finish completing their sentence and went over time.

The team responsible for the marking of the Handwriting Speed Test later confirmed to the project statistician that they had noticed the same pattern in many of the scripts: pupils had stopped writing once a sentence had been concluded and did not try to start a new one.

Given that it is not possible to know under which assumption each pupil completed the handwriting speed task, we have decided, deviating from the proposed Statistical Analysis Plan, to replace the original handwriting speed measure by number of times the sentence was written in full (HST_FS). HST_FS simply corresponds to the integer division of HST_RS by 35. The aggregation of letter counting into sentence counting implies a loss of granulation, but circumnavigates the problem of the original handwriting speed variable being ambiguously defined and measuring different outcomes. The distribution of the aggregated HST_FS variables, both general and in terms of randomisation groups, can be seen in a series of bar plots in appendices F1 and F2.

Sample size

The power calculations were performed with the calculations for a simple randomised design being adjusted for pre-post correlation and design effect using the Kish formula (Kish, 1965). All the calculations were performed assuming 80% power and alpha of 0.05. See Table 17 and Table 18 for a summary.

In the absence of a writing trial pilot, parameters for sample size were estimated using comparable EEF studies and materials. The 2015 EEF table of intra-cluster correlations (EEF, 2015) suggests a value of 0.109 for Key Stage 1 English in the North East and the 2013 pre-testing in EEF Evaluations paper (EEF, 2013) suggests a correlation of 0.73 between Key Stage 1 and Key Stage 2 English. The Grammar for Writing evaluation (Torgerson et al., 2014a) had a school-level ICC of 0.26 and the class-level ICC was 0.32. It used a predicted KS2 writing level as the baseline measure but the correlation was low at 0.54. The Calderdale Improving Writing Quality evaluation (Torgerson et al., 2014b) had a school-level ICC of only 0.04 for the extended writing task but this was based on only a subgroup of primary school children who went on to secondary schools within the trial. This trial also used a predicted KS2 writing level as the baseline and the correlation was also low at 0.35. Based on the research mentioned above, and to remain realistically conservative, we have adopted the values of 0.15 and 0.65 for the ICC and pre-post correlation, respectively.

According to a 2016 meta-analysis of handwriting interventions (Santangelo and Graham, 2016), handwriting instruction was associated with a large effect size of 0.84 on the quality of student writing.[8] On the other hand, recent studies provide evidence that EEF trials are underpowered (Sanders and Ni Chonaire, 2015 and Lortie-Forgues, 2017) so we considered an effect size of 0.15.

Considering an ICC of 0.15, a pre-post correlation of 0.65, and an effect size of 0.15 would result in trials requiring 140 (70 versus 70) schools. However, an efficacy trial on 140 schools would be too costly and entail a risk of diluting the intervention through limited delivery capacity. Taking into account these practical considerations we have settled for an effect size of 0.18 for the Year 2 trial and an effect size of 0.23 for the Year 5 trial,[9] both within what is expected from previous meta-analyses of handwriting interventions. These adjusted effect sizes required 100 schools (50 versus 50) for the Year 2 experiment and 370 Year 5 pupils (185 versus 185) for the Year 5 pupil-randomised experiment.

[8] It should be taken into account that the Santangelo and Graham meta-analysis included non-randomised designs.
[9] We reasonably expect the Year 5 intervention—a small group intervention targeted at selected pupils—to be more effective than the class-level Year 2 intervention.

Randomisation

Randomisation was carried out by a statistician at NFER using a full audit trail in R. The same statistician undertook the trial analysis aware of group allocation. Given the independence of the evaluation, we did not see it necessary to carry out the analysis blinded to randomisation. The randomisation code for both experiments can be found in Appendix S1.

The Year 2 experiment utilised school-level randomisation. To facilitate the delivery of training to school staff, four different training hubs were set up in different locations: Darlington, Leeds, Newcastle, and Sheffield. The assignment of schools to hubs took into account distance to the nearest or second nearest training hub to ensure that the training venues were reachable by the school staff undertaking the HHS training within a reasonable commuting time. Table 8, below, presents the assignment of the 103 schools taking part in the trial to training hubs.

Table 8: Distribution of participating schools in training hubs

Training hub	Number of schools	Percentage of schools	Maximum school-hub distance (miles)
Darlington	24	23%	24
Leeds	30	29%	13
Newcastle	28	27%	15
Sheffield	21	20%	17

The school-level blind randomisation for the Year 2 experiment took place in early August 2018,[10] after baseline testing being stratified by region (assigned training hub).[11] Schools were notified of the randomisation results after it took place. The stratification was introduced to prevent a clumping of intervention schools in certain regions that could impede delivery of the training. The school-level randomisation is described below, in Table 9.

Table 9: Year 2 experiment—school level randomisation regional strata

Region (training hub)	No. of schools in each arm		Total no. of schools
	HHS intervention	Control	
Darlington	12	12	24
Leeds	15	15	30
Newcastle	14	14	28
Sheffield	10	11	21
Total:	51	52	103

For sampling in the Year 2 experiment secondary outcomes analysis, we randomly selected five of the pupils taking part (per school) in the trial to be included in the secondary outcomes sample.[12] The trial was fully powered for the primary outcome only, but we acknowledge a discussion of power for secondary outcomes is sometimes useful, particularly as we were using a reduced sample here. Five was used as it is sufficient to allow follow-up of at least one pupil per school and is consistent with the idea that when estimating regression coefficients in a multilevel model, small

[10] Some of the participant schools only provided the results of the baseline assessment in the last weeks of July 2018, which led to the Year 2 randomisation taking place in early August 2018 instead of July 2018, as stated in the protocol.
[11] Baseline testing took place between the 20 June and 17 July 2018.
[12] In each school, the pupils were randomly selected from the Year 2 cohort taking part in the trial, with no stratification in terms of class or FSM eligibility being taken into account. If a school had fewer than five pupils taking part in the Year 2 trial, all the pupils were included in the secondary Year 2 sample.

cluster sizes are adequate.[13] This analysis was expected to have a higher MDES than for the primary outcome (approximately 0.25) but its inclusion was for verification of the comparative judgement method rather than intended to be fully powered. The sampling code can be found in Appendix S1.

We chose to select the secondary outcomes sample from the pupils present at pre-randomisation baseline assessment instead of selecting from the more restricted group of pupils with pre-test and post-test outcomes data. Although there was a risk of this procedure resulting in attrition, and consequent loss of power, we decided to adopt it in order to prevent biasing the analyses' results.[14]

Randomisation for the Year 5 experiment was conducted at the pupil level among participating schools. Six schools withdrew from the intervention group after the early August Year 2 school-level randomisation; 372 Year 5 pupils from the remaining 46 intervention schools were selected to take part in the trial and were randomised post-baseline testing in the second week of October 2018.[15] The randomisation was stratified by school and eligibility for FSM.[16]

The pupil-level randomisation breakdown by FSM eligibility status is described below in Table 10.

Table 10: Year 5 experiment pupil level randomisation—FSM eligible strata

FSM status	No. of pupils in each arm		Total no. of pupils
	HHS intervention	Control	
Not eligible	131	130	261
Eligible	54	56	110
Total	185	186	371

Statistical analysis

The primary and secondary analyses followed EEF 2018 guidelines for both the Year 2 and the Year 5 Experiments.

Primary outcome analyses

The primary analyses for both experiments were carried out on an intention-to-treat basis.

For Year 2, a multilevel random intercepts model with two levels (school and pupil) was used to account for cluster randomisation. The main analysis investigated whether the attendance of a class that received the HHS intervention had an effect on pupils' writing ability. This was determined by fitting a model with the dependent variable as writing ability post-intervention, as measured by the Writing Assessment Measure (WAM) comparative judgement true scores described above.

To control for prior writing ability, pupil-level WAM comparative judgement true scores assessed at baseline were included in the model as a covariate. The model also contained a dummy variable for region (training hub) to reflect the Year 2 stratified randomisation.

[13] See Snidjers et al., 2005, pp. 1570–1573.
[14] The rationale being that this ensured not only that bias was not introduced in the analysis but also that the primary and secondary analyses were performed under similar settings.
[15] Baseline testing took part between 19/11/2018 and 1/10/2018 and the randomisation on 9/10/2018. The baseline assessment scripts of one of the schools taking part on the trial were lost and the school was invited to retake the baseline test. The second assessment took part on 11/10/2018, four days before the random allocation result was disclosed to the school
[16] The FSM eligibility information was obtained directly from the schools during Year 5 data collection.

Since the Year 5 experiment was a multi-site efficacy trial, as per EEF 2018 guidelines we used a fixed effects single level model. The primary analysis determined if receiving the HHS intervention in a small group setting had an effect on the writing ability of pupils who struggle to produce fluent handwriting due to a deficit in fine motor skills. For this purpose we fitted a single-level regression model with the dependent variable as writing ability post-test as measured by WAM comparative judgement true scores.

Similarly to the Year 2 experiment, WAM comparative judgement true scores assessed at baseline were included in this model as a covariate. Dummy variables for school and FSM status were also included in the model to reflect the stratified Year 5 randomisation.

Secondary outcomes analyses

Analyses of secondary outcome 1—writing composition

For writing composition, we used an identical approach to the analyses of the primary outcome described above: we fitted a two-level model (pupil and school) with random intercepts to account for cluster randomisation for the Year 2 experiment, and a single level fixed effects model for the multi-site Year 5 Experiment. The dependent variable was writing composition post-intervention as measured by the Writing Assessment Measure (WAM) criterion referencing scores previously described. The baseline covariate for the models was the pre-test pupil-level WAM comparative judgement true scores; the randomisation stratifier-covariates were the same as for the ITT models.

Analyses of secondary outcome 2—handwriting speed

As described in the Outcome Measures section, it was necessary to aggregate the original handwriting speed letter-counting variable, HST_RS, into a sentence-counting ordinal variable, HST_FS. This represents a deviation from the SAP. The Poisson and negative binomial models prescribed in the SAP were no longer adequate to model the handwriting speed data and, as such, we had to resort to a different family of models: ordinal regression models. Ordinal regression models are suitable to model categorical data whose categories possess an implicit order.

The ordinal regression models we have adopted consider the effect of treatment allocation on children's handwriting speed by evaluating whether the probability of a pupil being classified in the higher or lower categories of the HST_FS variable measured post-intervention differs according to the pupil's randomisation group.

An ITT approach was taken for both experiments. The analyses were run by fitting models with post-intervention HST_FS as the dependent variable while controlling for handwriting speed at baseline by the inclusion of the categorical HST_FS variable derived from the HST_RS measured pre-intervention as a covariate. Covariates that reflected the stratified randomisations in the trial—school assigned training hub for the Year 2 experiment and school and reported FSM status indicator for the Year 5 experiment—were also included in the models.

To account for cluster randomisation on the Year 2 experiment, we have run two-level (pupil and school) random intercepts models, while in the context of the multi-site Year 5 experiment we ran single-level fixed effects models, following the EEF 2018 guidance.

For both Experiments we ran ordinal regression models with different link functions. We have selected as analysis models the ones with lower Akaike information criteria (AIC) and lower absolute values of log-likelihood (logLik). The selection of analysis models is displayed in Table 11 and Table 12 below. For the Year 2 experiment we selected a two-level random intercepts logit model and for the Year 5 Experiment a fixed effects cloglog model.

The models were run in R 3.6.1 (The R foundation for Statistical Computing, 2019) using the package 'ordinal' (Christensen, 2019).

To evaluate whether randomisation group (control vs intervention) was a significant parameter of the models, we have opted for a likelihood ratio test that can be obtained via the 'ordinal' package's ANOVA method. The default significance test provided by the 'ordinal' package relies on the Wald statistic and is not as accurate as the likelihood ratio test, which compares the model with a similar model with no randomisation group covariate under the null hypotheses that both models fit the data similarly.

Table 11: Selection of ordinal regression models for the Year 2 experiment

Model Type	Dependent variable	Model covariates	N obs	AIC	logLik
Two-level (pupil and school) random intercepts logit	HST_FS (post-intervention)	Control/intervention indicator; baseline: HST_FS (pre-intervention); dummy indicator: region	437	1349.21	-652.60
Two-level (pupil and school) random intercepts probit	HST_FS (post-intervention	Control/intervention indicator; baseline: HST_FS (pre-intervention); dummy indicator: region	437	1352.66	-654.33

Table 12: Selection of ordinal regression models for the Year 5 experiment

Model type	Dependent variable	Model covariates	N obs	AIC	logLik
Fixed effects cauchit	HST_FS (post-intervention	Control/intervention indicator; baseline: HST_FS (pre-intervention); dummy indicators: school and reported FSM status	328	1146.31	-505.16
Fixed effects cloglog	HST_FS (post-intervention	Control/intervention indicator; baseline: HST_FS (pre-intervention); dummy indicators: school and reported FSM status	328	1128.77	-496.39
Fixed effects logit	HST_FS (post-intervention)	Control/intervention indicator; baseline: HST_FS (pre-intervention); dummy indicators: school and reported FSM status	328	1137.93	-500.97
Fixed effects loglog	HST_FS (post-intervention)	Control/intervention indicator; baseline: HST_FS (pre-intervention); dummy indicators: school and reported FSM status	328	1157.71	-510.86
Fixed effects probit	HST_FS (post-intervention)	Control/intervention indicator; baseline: HST_FS (pre-intervention); dummy indicators: school and reported FSM status	328	1146.31	-505.16

Analysis in the presence of non-compliance

The compliance measures were designed by the evaluator, and commented upon by the developer, and consisted of three scales. The analysis in the presence of non-compliance is based on number and length of handwriting sessions delivered. This measure was included because it gives an indication of dosage (see Humphrey et al., 2016). It rates the number and length of all handwriting sessions delivered by the trained teacher. NFER provided teachers with a log in which they noted information on length, date, and so forth relating to all sessions delivered. For the Year 2 experiment, teachers recorded this information at class level. For the Year 5 experiment, teachers recorded this information at pupil level.

Each session was rated as 'not delivered', 'less than 30 minutes', '30 minutes', or 'more than 30 minutes'. Total intervention delivery time per school was calculated from this rating by summing each session length according to the rule described in Table 13:

Table 13: Rule for determining intervention delivery time ratings

Rating given by teacher per session	Time estimate to be used in measure
Not delivered	0
Less than 30 mins	20
30 mins	30
More than 30 mins	40

As this data reflects actual contact time between the children and the intervention, this formed the main Complier Average Causal Effect (CACE) analysis.

Taking into account the nature of the HHS intervention—school level for Year 2 and pupil level for Year 5—the Year 2 experiment's measure was a school-level dosage measure and the Year 5 experiment's a pupil-level dosage measure.

- Year 2 experiment school-level in the intervention group schools: average total length of HHS sessions delivered per class[17]
- Year 5 experiment pupil-level: total length of small group setting HHS sessions. This variable takes the value 0 for pupils in the control group, except two instances of contamination in which two control attended a total of 18 HHS sessions.

To evaluate if there was an association between the dosage of pupils with the HHS interventions and writing ability we treated the total intervention delivery time measure defined above as a pseudo-continuous dosage measure and adopted the instrumental variables approach (IV) prescribed by the EEF 2018 guidelines (Angrist and Imbens 1995, pp. 431–442).

In a systematic review of handwriting interventions, Hoy et al. (2011) concluded that interventions that included fewer than 20 practice sessions were ineffective. Based on the review's finding, we also defined the following dichotomous (Y/N) compliance variables of whether no fewer than 20 sessions were delivered/attended:

- **Year 2 experiment school-level**: all the Year 2 classes in the school had at least 20 out of 24 class-based HHS interventions delivered. This variable takes the value N for schools in the control group.
- **Year 5 experiment pupil level**: the pupil attended at least 20 out of 24 small group setting HHS interventions. This variable takes the value N for pupils in the control group.[18]

Although all of the 52 schools randomised into the control groups agreed to be assessed at follow up, seven withdrew from the trial without submitting trial logs and an additional two fully participating schools failed to submit them (see Figure 6: Participant flow diagram—the Year 2 experiment). Delivery logs completed at the class level by staff in the 43 intervention schools that submitted trial logs showed that the majority delivered between 20 and 24 HHS sessions to their Year 2 classes. Only four out of the 43 intervention schools that submitted logs delivered fewer than 20 sessions; their delivery of the intervention was therefore not compliant.

[17] The number of sessions delivered to Year 2 classes in each of the intervention schools was recorded by each classroom teacher delivering the HHS intervention in a fidelity/dosage log. For each of the intervention group schools, the Year 2 dosage measure was computed as $\frac{total\ length\ of\ sessions\ across\ teachers}{number\ of\ Y2\ classes}$.

[18] In the two instances of contagion, the control group pupils attended less than 20 interventions and were non-compliant.

We were able to collect compliance information for 168 of the 184 Year 5 intervention pupils who were not withdrawn from the trial; as one pupil left the school during the trial, this number reduced to 167. Of these, about four fifths (n = 134) received between 20 and 24 HHS sessions. A total of 33 (just under one fifth) received fewer than 20 sessions, which was not considered to be compliant. Compliance information was also provided for two control pupils who, according to the logs, attended HHS interventions.

We once again used instrumental variables (IV) approaches with group allocation as the instrumental variable, as suggested in Angrist and Imbens (1995, pp. 431–442), to enquire if there is an association between delivering or attending at least 20 sessions and the different trials' primary outcomes.

Missing data analysis

After evaluating to what extent data was missing and counting the number of complete cases, we proceeded to identify patterns of missingness in outcome variables.

As per the SAP, we were planning to investigate missingness just in terms of outcome variables. Given the design of the experiments—where only pupils who were assessed at baseline for the different outcomes were included in the trial—we were not expecting to find missing cases in the data corresponding to any of the covariates of the different models. However, once the trial data became available we came across missingness in terms of baseline WAM_CJ scores, which prompted us to deviate from the SAP. The reasons for the absence of baseline WAM_CJ scores are described and analysed in detail in the Impact Evaluation Results section below.

For the Year 2 experiment, we found instances of missing values of the primary outcome (WAM_CJ) for both the baseline and follow-up measurement adding up to a total of 433 cases (11% of the total number of cases). Since missingness in terms of baseline WAM_CJ is independent of the randomisation group allocated to the schools taking part in the trial (missing assessment data and impossibility to match scores to No More Marking), we have dismissed the possibility of bias being introduced in the analysis via this mechanism.

For the Year 5 experiment, we found instances of missing values of the primary outcome (WAM_CJ) for both the baseline and follow-up measurements adding up to a total of 58 cases (16% of the total number of cases).

We have identified missingness patterns in terms of outcome variables measured at follow up. After an exploratory analysis that hinted the presence of associations between the outcome variables and the covariates of the primary analysis model of the Year 2 experiment, the project statistician decided not carry out a Little's MCAR test (McKnight at al., 2007, pp. 93–94), whose null hypothesis is that data is missing completely at random (MCAR), and proceeded directly to investigate if the follow-up data was missing at random (MAR) for both experiments.

We have investigated missingness patterns by means of substantive models for the different outcome variables. For the Year 2 experiment, the outcome substantive model was a two-level (pupil and school) logistic model with baseline outcome, region, pupil gender, and randomisation group indicators as covariates; while for Year 5 we fitted a logistic regression model that included baseline outcome, school, reported FSM status, pupil gender, and randomisation group indicators as covariates.

Given that the Year 5 experiment had instances of attrition operating at both pupil and school level, we decided to supplement the missing data analysis with a re-run of the primary ITT analysis on a dataset with imputed baseline and follow-up missing values. The missing values were imputed using predictive mean matching, with five plausible values derived for each case. The imputation was done using the mice procedure contained in the R package mice (Van Buuren and Groothuis-Oudshoorn, 2011). The primary ITT model was re-run on five sets of imputed plausible values and the estimates for each model were then pooled into a single set of estimates and standard errors that was compared to the results of the original analysis.

Subgroup analyses

The primary outcome model for the Year 5 experiment has been modified for the FSM pupil analyses specified in the protocol. The corresponding analysis for the Year 2 experiment using the National Pupil Database (NPD) FSM6 data has now been performed and included in this revised report.

We have deviated from the standard EEF procedure of using 'ever FSM' and 'FSM6' as a deprivation indicator by using single point FSM eligibility instead for the Year 5 trial. This decision was motivated by the necessity to avoid collinearity in our regression models, since both ever FSM and FSM6 are known to be highly correlated with the FSM eligibility indicator already included as a stratifier in the Year 5 experiment models. A comparison between single point FSM and FSM6 from the NPD has been added to this revised report (see Table 35).

We approached the analyses in two distinct ways: we ran models with interaction terms (that is, models that include both the FSM indicator and the product of the FSM indicator and randomised group), and we ran separate primary outcome models on just the FSM-eligible pupils. Both approaches conform to the EEF 2018 guidelines (EEF, 2018).

The criteria for eligibility for the Year 5 trial include both pupils who are slow and effortful writers and pupils whose handwriting is illegible. To evaluate the differential impact of the small groups setting HHS intervention on slow and effortful writers versus children whose handwriting is illegible we also ran modified primary outcome models with interaction terms (handwriting speed raw scores at baseline and the product of the handwriting speed raw scores at baseline and randomised group were included in the model as covariates)[19].

Additional analyses and robustness checks

The design of this trial was specifically set up to track the theory of change: HST for improvement in handwriting, criterion reference marking for improvement in content, and comparative judgement for improvement in overall writing quality.

Estimation of effect sizes

As advised by the EEF 2018 guidelines, we report effect Hedges' g as effect sizes. These are calculated according to the formula:

$$\boldsymbol{g = \frac{\overline{o}_i - \overline{o}_c}{s^*}}$$

with $\boldsymbol{\overline{o}_i - \overline{o}_c}$ corresponding to the difference between the intervention and control group in terms of the mean value of the outcome being assessed, and $\boldsymbol{s^*}$ corresponding to the pooled standard deviation of the outcome.[20]

For both experiments, the numerator for the effect size calculation was the coefficient of the intervention group from the regression models (single level for Year 5, multilevel for Year 2). As prior ability is one of the covariates included in the models, we used unconditional total variance from the corresponding models without covariates as denominators. The effect size thus computed is equivalent to Hedges' g.

Ninety-five percent confidence intervals for each effect size were computed by converting the 95% confidence intervals for the intervention group coefficients using the same formula as the effect sizes themselves. The 95% confidence intervals were computed by means of the confidence interval estimation functions provided by the R-packages used to run the models.

[19] This subgroup analysis, not being pre-specified in the original protocol, will be reported as a post-hoc exploratory analysis.

[20] The pooled standard deviation is computed as $s^* = \sqrt{\frac{(N_i-1)s_{i+}^2(N_c-1)s_c^2}{N_i+N_c-2}}$ with N_i and N_C being the number of elements in the intervention and control groups, and s_i and s_c the standard deviations of the outcome measured in the intervention and control groups.

Estimation of ICC

For the Year 2 experiment, school-level ICCs were estimated from the variance of the random intercept and residual variance of the multilevel models by means of the formula:

$$ICC = \frac{\sigma^2_{intercepts}}{\sigma^2_{intercepts} + \sigma^2_{residuals}}$$

ICCs at baseline were computed considering random intercepts two-level (school and pupil) models with no covariates, and post-test ICCs were derived from the primary ITT model and secondary ITT model for the first secondary outcome described above (writing composition).

Longitudinal analysis

No longitudinal analysis has been planned but the data will be archived with the ONS for future use.

Implementation and process evaluation

Research methods

The purpose of the implementation and process evaluation (IPE) was to provide information on, and insights into, the delivery of Helping Handwriting Shine (HHS). The IPE investigated the following research questions:

- RQ1: Is fidelity to the intervention maintained?
- RQ2: How much does dosage differ across the sample?
- RQ3: To what extent do participants engage with the intervention?
- RQ4: What level and type of support does the developer provide to intervention schools?
- RQ5: What does 'business as usual' consist of?

A range of data collection methods were utilised in order to answer the IPE research questions. An overview of these are presented in Table 14 below.

Table 14: IPE methods overview

Research methods	Data collection methods	Participants/data sources	Data analysis methods	Research questions addressed	Implementation/ logic model relevance
Document analysis	Baseline proforma	Key contact (103 schools)	Frequency counts; inductive coding	RQ5	Context; usual practice
	Training attendance register	Training attendance register (46 schools)	Frequency counts	RQ1	Compliance; reach
	Fidelity and dosage logs	Year 2 staff logs (class and school level): – delivery log (51 Year 2 classes/43 schools)[21] – reflections log (31 schools) – monthly post-programme log (9 schools) Year 5 logs (pupil and school/group level): – delivery log (171 pupils/42 schools) – reflections log (34 schools) – monthly post-programme log (9 schools)	Frequency counts; inductive coding; thematic analysis	RQ1, RQ2, RQ3	Fidelity; dosage; adaptation; responsiveness; programme differentiation; cost
	'Business as usual' logs	Year 2 control school logs (class level 46) Year 5 control group logs (group level 35)		RQ1, RQ5,	Usual practice
Observations	Structured observations	Training workshops (4)	Thematic analysis	RQ1, RQ3, RQ4	Fidelity; adaptation; responsiveness;
Case studies (12; case study unit = school; analytical approach = methodological and participant triangulation)	Semi-structured interviews	Teachers (12) Support staff (12)	Inductive coding; thematic analysis	RQ1, RQ2, RQ3, RQ4, RQ5	Context; fidelity; dosage; adaptation; responsiveness; programme differentiation; quality; cost
	Semi-structured group interviews	Year 2 and Year 5 and pupils (36)	Inductive coding; thematic analysis	RQ1, RQ3	Context; responsiveness; adaptation
	Structured observations of intervention delivery	1 intervention session per case-study school (6)	Frequency counts; inductive coding; Cross-case analysis	RQ1, RQ3	Context; fidelity; adaptation; dosage; Responsiveness; programme differentiation; quality
Interviews	Structured interviews	Developers (3)	Inductive coding; thematic analysis	RQ1, RQ3, RQ4	Context

[21] Five schools completed logs for multiple Year 2 classes.

In order to collect data for this trial, we obtained the name and contact details of a key contact in each school (usually a senior leader) at the recruitment stage. Where schools were required to share personal data with us, we used the NFER secure portal to ensure the safe transfer of information. All research participants were assured confidentiality.

Further detail on each of the data collection methods is described below.

Document analysis

Baseline proforma for all schools

In June 2018, all schools recruited to the trial were asked to complete a short baseline proforma (an Excel spreadsheet; see Appendix O) at the pre-randomisation stage with details of their current/planned literacy support activity for Year 2 and Year 5 pupils over and beyond the normal curriculum. We sent an email containing a link to the NFER secure portal to the key contact in each school. School staff downloaded the proforma from the portal and once completed they uploaded it for the research team to access. Overall, 103 schools completed a baseline proforma. After randomisation, we were able to categorise baseline proforma responses to intervention schools (52) and control schools (51). See Figure 6 and Figure 7 for details.

Activity logs

Intervention schools were required to complete three logs (Excel spreadsheets; see Appendix P) to enable us to evaluate the fidelity, dosage, and implementation of HHS. These consisted of:

- Log 1: delivery log. A log during the flexible eight weeks of delivery (during the period November 2018 to March 2019)[22] recording what was delivered in each session. For Year 2, this was completed at the class level and for Year 5, this was completed for each intervention pupil.
- Log 2: reflections log. A one-off log after the eight-week intervention delivery period, in February 2019, reflecting on delivery.
- Log 3: post-programme log. A monthly log between the end of the eight-week intervention delivery period and the pupil endpoint testing (February–June 2019) reflecting on whether schools continued to use the programme.

Control schools were required to complete one log (an Excel spreadsheet; see Appendix P) to offer a comparison to business as usual literacy support. This consisted of:

- Business as usual log. A log during the period 1 November 2018 to 31January 2019 indicating the literacy support given to Year 2 classes in control schools and individual Year 5 target control pupils in intervention schools as part of the within-school trial design.

Activity logs were distributed to all schools via the key contact and in the same way as the baseline proforma. A guidance document was added to the NFER secure portal that described how to complete each log in detail. For intervention schools, further information about the logs was also provided by the developer at each training event using a PowerPoint presentation provided by the evaluator. Key contacts were provided with the details of a member of the NFER research team to contact if they experienced any issues with accessing or completing the logs.

Intervention logs

Log 1: delivery log
Between November 2018 and March 2019, schools were invited to complete a HHS ‘delivery log’ each week for their Year 2 class(es) (at class level) and their Year 5 intervention pupils (at pupil level). The Year 2 delivery log gathered delivery data on fidelity (for example, session length and frequency) during the eight weeks of programme delivery and whether key components of the intervention were delivered (for example, preparation, warm-up activities, writing activities, tailored levels, the model-plan-evaluate process, extension tasks, and gap tasks). Year 2 staff were also asked to log their ratings of class engagement during the intervention sessions and record any adaptations they made. Response option categories were provided in the log, where possible, to ease completion. The Year 5 delivery log was

[22] Schools could select their own start date for the eight-week intervention within this delivery period but were advised by the developer to administer it in two, four-week blocks.

similar to the Year 2 log but it was pre-populated with Year 5 pupil details so that information could be recorded at individual pupil level. We received feedback from some schools that completion of the logs was problematic for various reasons, including the time involved and difficulty using the spreadsheet software for some members of staff. Although the logs were based on a format that had been successfully used in other evaluations by the NFER team, they were not piloted in the format designed for this trial, which may have led to their simplification and/or curtailment. Email and telephone reminders were used to increase response rates and where schools had multiple Year 2 classes they were informed that only one Year 2 log was required in order to reduce burden. Overall, 51 Year 2 delivery logs were completed (at class level) from 43 schools (96% response rate) and Year 5 delivery logs (for 168 pupils) were completed from 42 schools (95% response rate).

Log 2: reflections log

Following the eight-week intervention period, in February 2019, all participating Year 2 and Year 5 staff were invited to complete a 'post-intervention reflections log'. They were required to record how frequently they made adaptations to the programme, the approximate costs incurred, the time spent per week on the intervention outside of teaching/delivery (for example, for planning or preparation), any issues or challenges faced during delivery and any reflections on the programme more generally. Response option categories were provided where possible in the log to ease completion. Email and phone reminders were used to increase response rates. A total of 31 schools completed a Year 2 post-intervention reflections log (69% response rate) and 34 schools completed a Year 5 log (75% response rate).

Log 3: post-programme log

To understand any further literacy support in intervention schools beyond the eight-week intervention delivery period, all Year 2 and Year 5 staff were also invited to complete a 'monthly post-programme log' from February 2019 until the pupil endpoint testing in June and July 2019. The post-programme log asked school staff to record the extent to which HHS had been embedded through the continued use of activities, techniques, or gap tasks. The monthly post-programme log also required staff to record any support they had requested or received from the developer and whether they had participated in any other external support or CPD for handwriting specifically, or literacy more generally, during this time. We received an increasing amount of feedback from schools that completion of the logs was becoming problematic. There was a risk of overburdening schools during a time when we were also trying to arrange administration of the pupil endpoint testing with schools and therefore it was decided that a reminder strategy to increase response rates should not be implemented. In total, nine schools returned Year 2 monthly post-programme logs and nine schools returned Year 5 post-programme logs (response rates of 20%).[23] In order to ensure we still captured data on the support schools received following the intervention period and the extent to which HHS had been embedded, these topic areas were included in the case-study telephone interviews that took place in the summer term 2019 (see further details below).

Business as usual log

In January 2019, NFER sent an email to the key contact in control schools asking Year 2 staff to complete a business as usual log about the literacy support given to their Year 2 class. Due to the within-school trial design, Year 5 control pupils came from intervention schools, key contacts in these schools were emailed to ask a member of Year 5 staff to complete a business as usual log about Year 5 control group pupils. The main purpose of the these logs was to enable us to understand any differences in the support received by control pupils and intervention pupils and to identify any potential contamination between intervention and target control pupils in Year 5. We specified that the logs should be completed during the period 1 November 2018 to 31 January 2019 (to align with the intervention delivery period). The log required respondents to record any allocated time specifically for handwriting practice in class, the details of any specific handwriting practice sessions, and whether any specific handwriting interventions or literacy programmes were delivered during this period. Email and telephone reminders were used to increase response rates. A total of 46 Year 2 (class level) business as usual logs were received with a response rate of 90% and 35 Year 5 (pupil level) business as usual logs were received with an 80% response rate.

Training attendance register

The HHS team offered training to all intervention schools in October and November 2018. This training was delivered as a one-day session in four regional 'hub' locations—Leeds, Sheffield, Darlington, and Newcastle (two were delivered per area, with eight training sessions delivered in total). A minimum of two members of staff (preferably a Year 2 teacher,

[23] These were not all the same schools.

a Year 5 teacher or teaching assistant, and a senior leader) were required to attend from each school. The developer collected a record of attendance at each session; this was analysed to measure compliance.

Observations of training workshops and follow-up CPD

During October and November 2018, an NFER researcher attended one HHS training event in each of the four regional hub locations. These observations allowed us to better understand the programme and how it should be delivered (including permissible adaptations) and to observe the levels of engagement and interaction amongst attendees. The same member of the research team attended all four training events in order to assess the consistency of delivery and any variation in participant engagement.

The developer also delivered two half-day follow-up CPD sessions in each hub location (morning and afternoon) during February and March 2019. The purpose of this session was to help school staff transition from the formal eight-week intervention to a way of embedding recommended practices (that is, gap tasks) into their regular teaching. The follow-up CPD was not part of the original theory of change or protocol; this was added at a later stage to support schools and encourage embedding during the period between the end of the formal intervention and the endpoint testing. As such, school attendance was not pre-identified as a specific compliance measure for this trial and attendance data was not collected. A member of the research team did, however, observe one of the half-day CPD sessions. The purpose of this observation was to gain an understanding of the content of the session and the nature of the support available from the developer at this stage of the intervention.

Case studies

In November 2018, after intervention schools had participated in the HHS training, we emailed the key contact in each school to ask if they would be willing to take part in the IPE as a case-study school.Of the schools that agreed to take part, a sample of 12 case-study schools were chosen, ensuring a range of geographical areas (defined by training hub locations and local authority areas) and a range of Year 2 and Year 5 class or group sizes were represented. These criteria were chosen in order to select a range of schools from across the sample, rather than purely random. Half of these schools were selected for case-study visits and half were allocated to take part in lighter-touch telephone interviews.

Table 15: Summary of case study details

Case Study ID	Method	Training Hub location	LA
CS1	School visit	Leeds	Bradford
CS2	School visit	Leeds	Bradford
CS3	School visit	Newcastle	Durham
CS4	School visit	Darlington	Darlington
CS5	School visit	Sheffield	Sheffield
CS6	School visit	Sheffield	Sheffield
CS7	Telephone interviews	Newcastle	Gateshead
CS8	Telephone interviews	Newcastle	South Tyneside
CS9	Telephone interviews	Newcastle	Durham
CS10	Telephone interviews	Darlington	Darlington
CS11	Telephone interviews	Sheffield	Sheffield
CS12	Telephone interviews	Leeds	Leeds

The purpose of the case studies was to understand how different schools were implementing the intervention in different contexts and situations, to gain face-to-face feedback from teachers and pupils, and to observe implementation fidelity. In January and February 2019, NFER researchers visited six case-study schools to observe intervention delivery during the trial period. All of the case-study schools had been delivering HHS for at least four weeks by the time of the visit. At

least one observation of a HHS session was carried out at each school. We recorded information on the number of participating pupils and staff, details of the delivery space, whether it was a whole-class activity or small group activity, how it was delivered (and any adaptations made), levels of pupil engagement, and whether any practical problems were encountered. The observations were also complemented by face-to-face interviews with two members of staff (one undertaking Year 2 delivery and one undertaking Year 5 delivery). All staff participating in the interviews were those who had direct experience of delivering HHS and could therefore give feedback on the practicalities of running the programme in their school and the school or class context. The interviews were semi-structured and lasted up to 30 minutes. As an introduction, we explained the purpose of the project, identified the topics we were going to cover and guaranteed confidentiality. The interview topics covered:

- preplanning and foundations—the background to the schools' involvement with HHS, the level of need, readiness and capacity for developing handwriting in the school, and its fit with wider school strategies;
- implementation support system—the training and support that was accessed from the University of Leeds team and how is it was perceived;
- implementation environment—how the intervention fits with, or differs from, the school's usual practice, any senior leader support, and any barriers to delivery or key features of effectiveness;
- implementation factors—who delivers HHS and how is it delivered (for example, fidelity regarding timing and dosage), permissible tailoring, and any aspects of the programme or materials that were revised or adapted; and
- intervention impacts—impacts of the intervention on pupils and staff (and, where interviewees were senior leaders or literacy coordinators, impacts on the wider school).

Pupil discussion groups were carried out during each case-study school visit with Year 2 or Year 5 pupils (as practical and appropriate)[24] in order to inform our understanding of that particular situation. These discussions focused on the delivery of the intervention, whether they liked it or not, and whether they felt it helped their handwriting. Teachers and TAs were asked to select pupils from their class or intervention group to take part in the discussion groups. Strategies that the IPE researchers felt would enhance their ability to talk with young (especially Year 2) pupils were utilised, including using flashcards to help pupils articulate their views. A total of 23 Year 2 pupils and 13 Year 5 pupils took part.

Between June and July 2019, after the formal intervention delivery period had finished and when schools were in the 'embedding' phase, we conducted telephone interviews with two members of school staff (one responsible for delivering to Year 2 and another delivering to Year 5) in a further six schools. The telephone interviews were semi-structured, lasted up to 30 minutes, and covered the same range of topics as outlined above with some additional areas of questioning to explore the post-intervention period, including any further support received by schools from the developer and the extent of embedding activities.

Developer interviews

In March 2019, we also conducted face-to-face interviews with three members of the University of Leeds team to explore perceptions of the effectiveness of the training and follow-up support, the intervention design and its implementation, as well as any challenges encountered and how these were overcome.

Rationale for the data collection methods utilised

Researchers chose the range of data collection methods outlined above as they offer both breadth and depth to the implementation and process evaluation. They provide good coverage across the IPE dimensions and ensure each of the IPE research questions are adequately evidenced. The theory of change was used to help prioritise data collection on the key features of the 'theory of change' for the intervention. We considered that the methods were appropriate for this trial because they enabled us to examine how schools delivered the programme, both at the whole-class level (for Year 2s) and in small groups (for Year 5s) during the intervention period, while also collecting data from schools after the intervention period to understand whether, or how, they had embedded HHS into their usual practice. The logs provide an efficient way of measuring compliance and fidelity and inform our understanding of usual practice at a granular level across a large number of schools. However, as these are self-reported measures, we also sought to gain

[24] In some schools, the Year 5 pupils in the intervention group had SEN and school staff felt that their specific needs and conditions meant that participation in research activities would not be appropriate at that particular time. We were unable to gain consent by proxy from schools in such cases and therefore just the Year 2 pupils were interviewed.

observational evidence of how well delivery worked, what constraints and delivery challenges schools encountered, and how they addressed them. Combined with additional qualitative information gained from interviews with school staff and pupils within different contexts, these data collection sources yield valuable information on the feasibility of the future roll-out of the intervention and further insight into the impact of HHS.

The NFER research team developed all of the data collection instruments (the logs and interview and observation schedules). The delivery log was designed by the evaluator and commented upon by the developer to ensure it captured data on the key components of the intervention. The research instruments were developed to meet the specific requirements of the trial, although they follow a similar format to others used in EEF trials conducted by NFER. The NFER research team collected all data from schools via logs, interviews, and observations. The University of Leeds collected and collated attendance register data. The register contained pre-specified fields for the purposes of the trial as specified by the research team. This information was shared securely with NFER. The research team checked the reliability of attendance data against the attendance records they collected at the observed training days.

Analysis

Separate NFER teams conducted the outcomes analysis and process analysis in this trial. We initially analysed and reported process data prior to knowing trial outcomes to avoid biased interpretation. The theory of change was used to develop a framework for IPE analysis in order to explore the main aspects of the intervention and relationships between them. In our overarching analysis of proforma data, log data, interview data, and observation data we examined the responses to common topics. We identified the numeric values from the logs for each column or activity (see Appendix P for logs) and cross-referenced the responses with the interview and observation data (which was coded using an inductive approach and analysed thematically) to present a holistic quantitative and qualitative picture of intervention delivery. This approach enabled us to provide a broad and in-depth evidence-based account and assessment of the delivery of the intervention. Quantitative analysis undertaken by the IPE research team was quality assured by a project statistician.

Costs

Intervention schools were asked to contribute £500 in order to receive the intervention; this was 25% of the actual cost of the intervention. This amount covered staff training sessions, training resources, an intervention handbook, pupil resource booklets, and a resource pack (including stickers, playdough, etc.). Intervention schools made payment at the point of booking on to the training, and this was dealt with entirely by the University of Leeds.

In addition, intervention schools were required to cover the costs of:

- supply cover for staff attending training;
- travel costs associated with attending training; and
- teacher time commitments.

Control schools received £500, paid to them by the University of Leeds after completion of the post tests, and requiring them to have completed all necessary elements of the trial including provision of pupil data and test data, fidelity monitoring logs, proformas covering what business as usual consists of, and anything else required for the evaluation.

Information was collected about the cost of the intervention as it was delivered in the evaluation, and about what it would cost a school to deliver HHS. As the programme is partially funded for intervention schools by the EEF and the University of Leeds, further cost information was also sought from them. The developer also provided us with information on the total number of individuals (from each intervention school) that attended the training. The number of HHS sessions for each intervention pupil was collected via the delivery log, which was sent to us by intervention schools at the end of the delivery period. We asked all the intervention schools to provide their cost data by including cost-related questions in the post-intervention reflections logs and in interviews with school staff. We explored direct costs, marginal costs—including subsistence and travel costs, and costs for resources. All of the above were used to calculate and report on the costs of delivering HHS. We did not collect business as usual cost data as this was deemed to be too onerous given the potentially wide range of literacy support activities and interventions that schools would be delivering to control pupils.

Costs were calculated as a cost per pupil from the school's perspective, as if schools were paying for the intervention, based on marginal financial costs.

Staff time

In the reflection logs and in interviews, we asked intervention school staff to provide information on the average amount of time they invested in preparing for each HHS session (not including engaging in evaluation activities, such as providing data, completing logs, or interviews). The time estimate was then calculated using the same methodology as the financial cost estimate.

Timeline

Table **16** shows the timeline of the complete evaluation.

Table 16: Timeline

Dates	Activity	Staff responsible / leading
January–March 2018	Write protocol	Gemma Stone
March–May 2018	Recruitment Developer interviews	Kathryn Hurd, Gemma Stone Gemma Stone
June 2018	Baseline assessment and data collection for Year 2 experiment Baseline data collection for Year 5 experiment	Gemma Stone, Kathryn Hurd
July 2018	Randomisation for Year 2 experiment	Joana Andrade Gemma Stone
September 2018	Baseline assessment for Year 5 experiment	Gemma Stone, Kathryn Hurd
September 2018	Randomisation for Year 5 experiment	Joana Andrade Gemma Stone
October 2018	Intervention schools attend training sessions NFER observes training sessions	Gemma Stone Kerry Martin
November 2018	Intervention schools begin intervention delivery Intervention schools complete weekly logs Control schools continue BAU	(Schools)
December 2018–January 2019	Intervention schools continue intervention delivery Intervention schools continue to complete weekly logs Case study visits	(Schools) Kerry Martin
February 2019	Some intervention schools continue intervention delivery Some intervention schools continue to complete weekly logs Control schools complete BAU logs Case-study visits Reflections log Monthly post-programme log Observe follow-up CPD	(Schools) Kerry Martin
March 2019	Some intervention schools continue intervention delivery Some intervention schools continue to complete weekly logs Monthly post-programme log Developer Interviews	(Schools) Gemma Stone
April 2019	Monthly post-programme log	(Schools)
May 2019	Monthly post-programme log	(Schools)

Dates	Activity	Staff responsible / leading
June 2019	Monthly post-programme log Follow up testing for Year 2 and Year 5 experiment Case-study telephone interviews	(Schools) Gemma Stone Kathryn Hurd Kerry Martin
July 2019	Follow up testing for Year 2 and Year 5 experiment Case-study telephone interviews Marking/judging	Frances Brill Gemma Stone Kathryn Hurd Kerry Martin
August–September 2019	Marking/judging Data cleaning and compilation	Frances Brill Gemma Stone
October 2019–February 2020	Analysis and reporting	Joana Andrade Gemma Stone
March 2022	NPD analysis of FSM6 data	Joana Andrade
September 2022	Publication of report with NPD analysis	Ben Styles

Impact evaluation results

Participant flow including losses and exclusions

Figure 6: Participant flow diagram—the Year 2 experiment

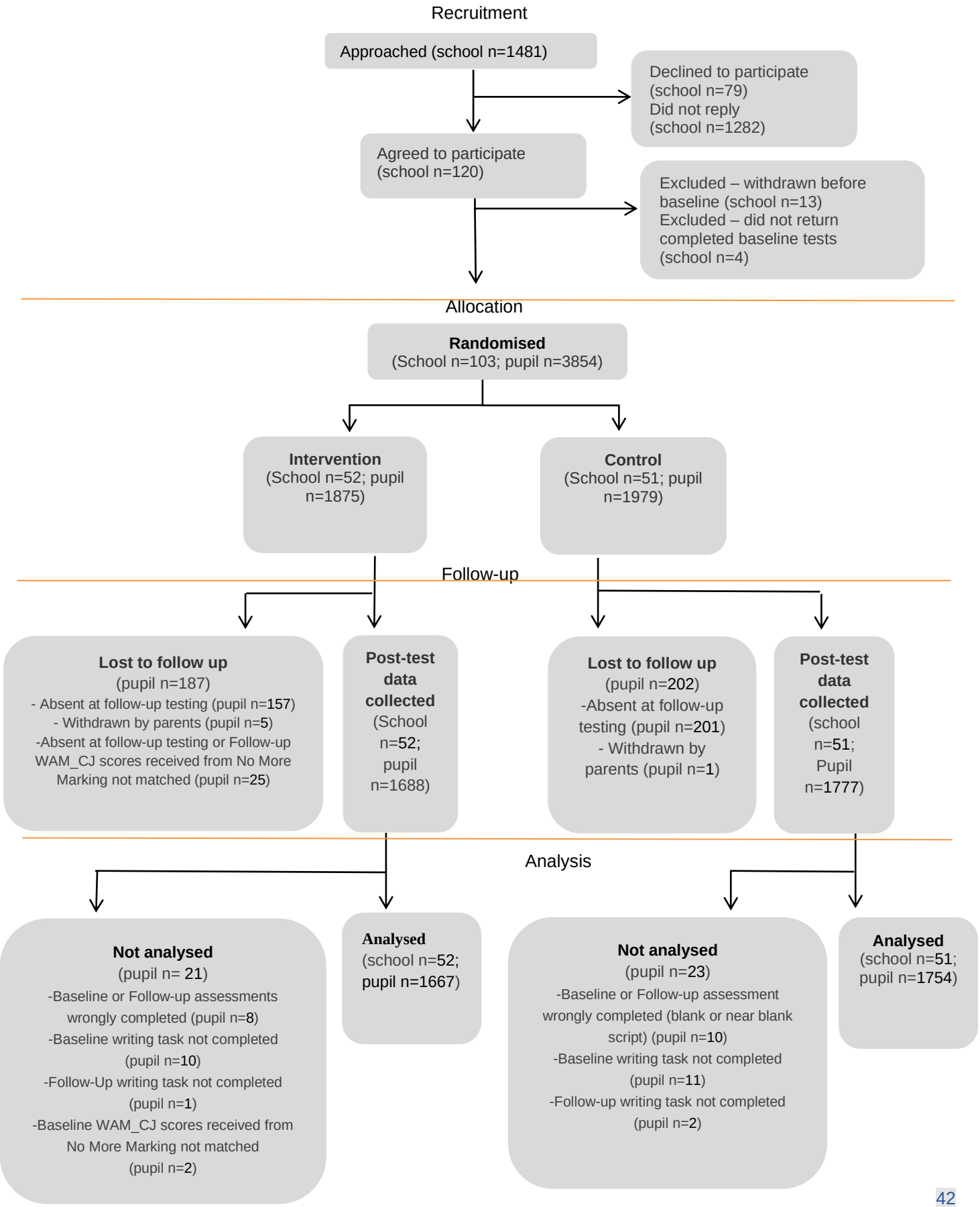


Figure 7: Participant flow diagram—Year 5 experiment

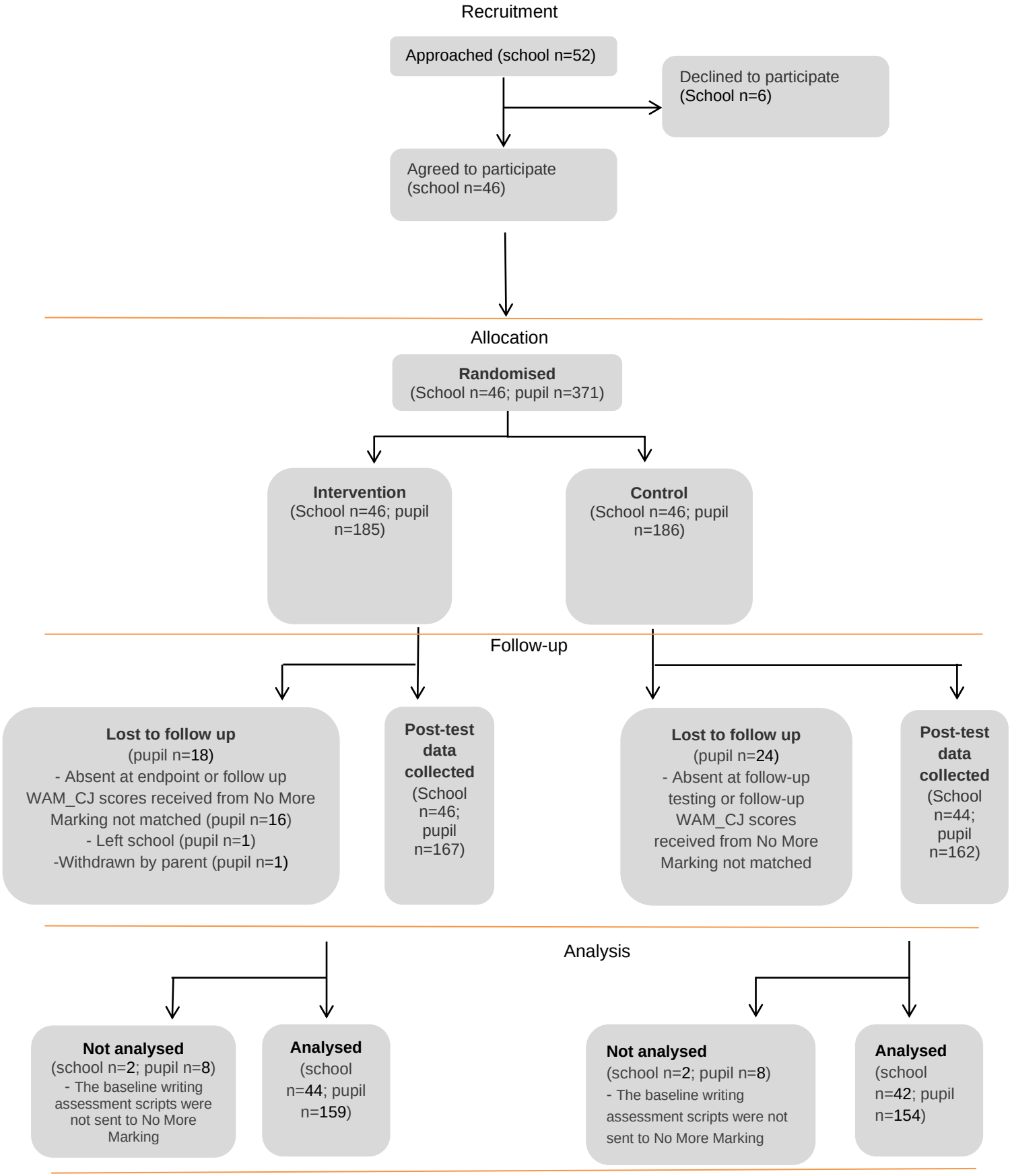


Table 17: Minimum detectable effect size at different stages—Year 2 experiment

		Protocol		Randomisation		Analysis	
		Overall	FSM	Overall	FSM[25]	Overall	FSM
MDES		0.18	0.21	0.18	0.23	0.16	0.20
Pre-test/post-test correlations	Level 1 (pupil)	0.65	0.65	0.65	0.65	0.66	0.66
Intracluster correlations (ICCs)	Level 2 (school)	0.15	0.15	0.15	0.17	0.12	0.10
Alpha		0.05	0.05	0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8	0.8	0.8
One-sided or two-sided?		Two-sided	Two-sided	Two-sided	Two-sided	Two-sided	Two-sided
Average cluster size		37	11	37	10	33	8
Number of schools	Intervention	50	50	52	51	52	51
	Control	50	50	51	47	51	46
	Total:	100	100	103	98	103	97
Number of pupils	Intervention	1850	574	1875	529	1667	437
	Control	1850	574	1979	413	1754	337
	Total:	3700	1148	3854	942	3421	774

[25] Parameters for FSM6 analysis have been added in this revised report.

Table 18: Minimum detectable effect size at different stages—Year 5 experiment

		Protocol		Randomisation		Analysis	
		Overall	FSM	Overall	FSM	Overall	FSM
MDES		0.23	0.23 to 0.41*	0.23	0.41	0.24	0.44
Pre-test/post-test correlations	Level 1 (pupil)	0.65	0.65	0.65	0.65	0.68	0.67
Alpha		0.05	0.05	0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8	0.8	0.8
One-sided or two-sided?		Two-sided	Two-sided	Two-sided	Two-sided	Two-sided	Two-sided
Number of schools	Total:	50	50	46	39	44	38
Number of pupils	Intervention	185	54 to 185*	185	54	159	42
	Control	185	54 to 185*	186	56	154	49
	Total:	370	108 to 370*	371	110	313	91

* Based on the assumption that the probability of an FSM-eligible pupil being eligible for the trial is identical to the overall probability of being eligible, the expected number eligible per school is 2.16. Under the assumption that all eligible pupils are FSM-eligible, the expected number per school was 7.4 as per the main sample size calculation. The true value lies somewhere in between.

For the Year 2 experiment (Table 17), we used FSM6 eligibility as this ultimately comes from the NPD. For the Year 5 experiment (Table 18), we used FSM eligibility since this was a randomisation stratifier obtained directly from the school. These tables indicate that parameter estimates were in line with those estimated. The number of schools in the Year 2 experiment was slightly higher than predicted (103 as compared to 100) and the ICC slightly lower, so the MDES at analysis for that experiment reduced to 0.16. However, numbers of pupils reaching analysis stage were slightly down on those predicted for Year 5. This led to a slight increase in MDES from 0.23 to 0.24.

Attrition

Treatment attrition occurred at both pupil and school level; the percentages of treatment attrition can be seen in Table 19 and Table 20. Treatment attrition was mainly due to school-level withdrawals and, in the case of the Year 2 experiment, affected intervention schools more severely than control schools, for which it was negligible.

Although all the randomised schools that took part in both experiments agreed to have their pupils assessed at follow-up, seven schools withdrew from the Year 2 intervention, accounting for a total of 286 pupils, and two schools withdrew from the Year 5 intervention, accounting for a total of 16 pupils. At the pupil level, we have only recorded a total of eight withdrawals from the trial (six Year 2 pupils and one Year 5 pupil were withdrawn by their parents or guardians, and one Year 5 pupil left the school during the trial).

Table 19: Pupil-level treatment attrition from Year 2 experiment—primary outcome

		Intervention	Control	Total
Number of pupils	Randomised	1875	1979	3854
Pupil attrition (trial)	Withdrawn	291	1	292
	Percentage	15%	0.05%	8%

Table 20: Pupil-level treatment attrition from the Year 5 experiment—primary outcome

		Intervention	Control	Total
Number of pupils	Randomised	185	186	371
Pupil attrition (trial)	Withdrawn	10	8	18
	Percentage	5%	4%	5%

Reasons for withdrawal from the intervention at a school level were as to be expected, such as concerns over staff time and the requirements of the school in terms of providing cover for the training and time out of lessons. There were also some more specific issues:

- Some schools indicated frustration at not having enough information at point of recruitment about the exact nature of the intervention—that they were not made fully aware of the contents of the programme until they had received the training. One school withdrew at this stage on the basis that the techniques were not new to it and that the investment (in terms of staff time and money) was not justified.
- Intervention schools were required to pay for the intervention in this trial, at a cost of £500, which represents 25% of the value of the intervention. Some schools objected to this, though only one school cited this as a reason for withdrawing.
- Specifically for the Year 5 experiment, teachers expressed consternation over the method of sampling and randomisation within the Year 5 classes. No issues were raised over the staff selecting the pupils who would most benefit from the intervention, but it would appear that staff were unhappy once these pupils were randomly allocated to control or intervention. Staff reported via the process evaluation that the resulting intervention groups often comprised the 'wrong' pupils, were occasionally unbalanced, and often too small.

All the randomised schools that took part in the trial agreed to follow-up assessment, even the ones that withdrew from the intervention. Measurement attrition operated only at pupil level for the Year 2 experiment and at pupil and school level for the Year 5 experiment. Pupil-level attrition was mainly due either to pupils being absent on the date of follow-up testing (358 to 383 Year 2 pupils and up to 40 Year 5 pupils) or problems with assessment data collection and processing (44 to 69 Year 2 pupils and up to 40 Year 5 pupils). The baseline assessment scripts of two Year 5 participating schools (16 pupils) were not sent to No More Marking and, therefore, no baseline data was available to be included in the analysis. The numbers of pupils lost to the analysis due to factors like parental withdrawal or pupils having left the school during the trial was comparatively much smaller (six Year 2 pupils and two Year 5 pupils). Recorded reasons for absence at follow-up testing included sickness on the day of testing. See Figure 6, pupil flow diagram, for further details of attrition.

Table 21: Pupil-level measurement attrition from the Year 2 experiment—primary outcome

			Intervention	Control	Total
Number of pupils	Randomised	1875	1979	3854	
	Analysed	1667	1754	3421	
Pupil attrition (from randomisation to analysis)	Number	208	225	433	
	Percentage	11%	11%	11%	

Table 22: Pupil-level measurement attrition from the Year 5 experiment—primary outcome

			Intervention	Control	Total
Number of pupils	Randomised	185	186	371	
	Analysed	159	154	313	
Pupil attrition (from randomisation to analysis)	Number	26	32	58	
	Percentage	14%	17%	16%	

Pupil and school characteristics

As specified in the SAP, we have also evaluated whether the distribution of schools per region (training hub) differs significantly when considering their randomisation group since all the schools that took part in the Year 2 experiment were present at both baseline and end-point, meaning that the randomised and analysed groups coincide. We have performed a Chi-squared test and obtained a p-value of 0.99 , and as such we did not reject the null hypothesis that the regional distribution of schools does not differ between the control and intervention groups.

We have also tested if the groups differed in terms of the following school-level indicators: school governance, school type, rural-urban, Ofsted rating, average proportion of 'ever FSM' pupils, and reading performance at KS2.[26] With the exception of the school governance indicator where we found a statistically significant difference between the treatment groups ($p = 0.04$ for a Chi-squared test), we failed to reject the hypothesis that the two groups were significantly different. Note that when carrying out multiple tests on randomised groups we would expect there to be Type I errors such as this. The results of the comparison between the Year 2 experiment's randomised groups in terms of school-level characteristics can be seen in Table 23.

Also, as per the SAP, we have tested if the mean baseline scores of the primary outcome differed between control and intervention schools (as randomised and analysed) by fitting two-level (school and pupil) random intercepts models. We have obtained, for a Satterthwaite degrees of freedom test, a p-value of 0.28 (randomised) and 0.55 (analysed), and so we did not reject the null hypothesis that the two groups have the same average baseline WAM_CJ scores in both cases. The results of the analysis of pupil-level characateristics of the randomisation groups for the Year 2 experiment are displayed in Table 24, below.

[26] Although the pupils participating in the Year 2 experiment were in KS1, we have chosen to report on schools' average reading performance at KS2 as a proxy for reading performace at KS1. All the schools participating in the trial included both KS1 (Year 2) and KS2 (Year 5) classes, but also this measure is of relevance for the evaluation of the results of the Year 5 experiment.

Table 23: Year 2 experiment—school-level baseline characteristics of the randomisation groups

School-level (categorical)		Intervention group		Control group		
		n/N[27] (missing)	Count (%)	n/N (missing)	Count (%)	p-value*
School governance Academy or Free School Maintained		22/52 (0) 30/52 (0)	42.3% 57.7%	11/51(0) 40/51 (0)	21.6% 78.4%	0.04
School type Primary		52/52 (0)	100%	51/51 (0)	100%	1.00
Urban or rural Urban Rural		8/52 (0) 42/52 (0)	16.0% 84.0%	10/51 (0) 41/51 (0)	19.6% 80.4%	0.83
Ofsted rating Outstanding Good Requires improvements Inadequate		9 /52 (0) 37/52 (0) 3/52 (0) 3/52 (0)	17.3% 71.2% 5.8% 5.8%	15/51 (0) 29/51 (0) 7/51 (0) 0/51 (0)	29.4% 56.9% 13.7% 0	0.07
Region Darlington Leeds Newcastle Sheffield		12/52 (0) 15/52 (0) 14/52 (0) 11/52 (0)	23% 29% 27% 21%	12/51 (0) 15/51 (0) 14/51 (0) 10/51 (0)	24% 29% 28% 20%	0.99
School-level (continuous)	National-level** mean	n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	p-value*
% FSM Ever (2016/2017)	22.51%	49 (3)	28.23% (14.62)	51 (0)	26.46% (16.52)	0.57
Average KS2 reading point score (2017)	104.27	51 (1)	103.33 (2.94)	49 (2)	104.04 (3.40)	0.27

Source: NFER Register of Schools (2018)

* Chi-squared tests, except for Ofsted rating where a Fisher's exact test was performed due to small cell sizes.

** All English primary schools that teach both Year 2 and Year 5.

*** T-tests following a Levene's test.

[27] n refers to the sample size of a specific group (for example, intervention schools) while N refers to the total sample size (for example, all schools in the trial).

Table 24: the Year 2 experiment—baseline pupil-level characteristics of groups as randomised and analysed

	Intervention group		Control group			
	n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	Effect size	p-value
Baseline WAM_CJ (randomised)	1855/1875 (20)	-0.168 (2.63)	1958/1979 (21)	0.159 (2.60)	-0.125	0.28
Baseline WAM_CJ (analysed)	1667/1875 (208)	-0.047 (2.56)	1754/1979 (225)	0.222 (2.61)	-0.104	0.55

The schools that took part in the Year 5 experiment were recruited from the group of intervention schools of the Year 2 experiment. As such, the distributions of school-level indicators for the Year 5 experiment are similar to those of the Year 2 experiment's intervention group and can be consulted in Table 23.

Once again, and as per the SAP, we have enquired if the prevalence of FSM eligibility, as reported by schools, differed significantly between the randomisation groups of the Year 5 experiment (as randomised and analysed). We have performed Chi-squared tests and obtained p-values of 0.94 (randomised) and 0.35 (analysed) and so we did not reject the null-hypothesis that the prevalence of FSM eligibility is equal in both groups, both at randomisation and at analysis. Following the SAP, we also ran t-tests28 to evaluate if mean baseline scores of the primary outcome differed between control and intervention pupils (as analysed). The t-tests had associated p-values of 0.47 (randomised) and 0.31 (analysed), and, as such, we did not reject the null hypothesis that the two groups have the same average baseline WAM_CJ, for both the randomisation and the analysis stage. The results of the pupil-level comparisons between the Year 5 experiment's groups are displayed in Table 25 (randomised) and in Table 26 (analysed).

Table 25: Year 5 experiment—baseline pupil-level characteristics of groups as randomised

FSM status (reported)	Intervention group		Control group			
	n/N[29] (missing)	Count (%)	n/N (missing)	Count (%)	p-value	
FSM_eligible	54/185 (0)	29%	56/186 (0)	30%	0.94	
Non_eligible	131/185 (0)	71%	130/186 (0)	70%		
Baseline	n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	p-value	Effect size
WAM_CJ	175/185 (10)	0.087 (2.10)	178/186 (8)	-0.083 (2.32)	0.47	0.08

[28] In preparation for the t-tests, we tested if the control and intervention groups had the same variance by means of a Levene's test. It did not reject the hypothesis of both groups having the same variance—p-value = (randomised); p-value=0.30 (analysed)—and, as such, the t-test were run assuming equal variances.
[29] n refers to the sample size of a specific group (for example, intervention schools) while N refers to the total sample size (for example, all schools in the trial).

Table 26: Year 5 experiment—baseline pupil-level characteristics of groups as analysed

FSM status (reported)	Intervention group		Control group			
	n/N[30] (missing)	Count (%)	n/N (missing)	Count (%)	p-value	
FSM_eligible	42/185 (12)	26%	49/186 (7)	32%	0.35	
Non_eligible	117/185 (14)	74%	105/185 (25)	68%		
Baseline	n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)	p-value	Effect size
WAM_CJ	159/185 (26)	0.132 (2.09)	154/186 (32)	-0.121 (2.34)	0.31	0.11

[30] n refers to the sample size of a specific group (for example, intervention schools) while N refers to the total sample size (for example, all schools in the trial).

Outcomes and analysis

Primary analysis

Histograms describing the distributions of the comparative judgement scores (WAM_CJ) for both baseline and follow-up can be found in Appendices D1 and D2. Each histogram includes a normal curve with the same mean and standard deviation of the underlying WAM_CJ distribution. Although the baseline and follow-up data of both experiments do display some deviation from normality, the QQ plots of fitted versus observed residuals (also in Appendices D1 and D2) suggest that the analysis models fitted the data adequately.

Displayed in Table 27 are the findings for the primary analysis of both experiments. As described in the methods section, we have run regression models: a two-level random intercepts model for the Year 2 experiment and a single-level fixed effects model for the Year 5 experiment, and presented their results as effect sizes (Hedges' g). The effect size for the Year 2 primary analysis was -0.02 (-0.13, 0.1) and for the Year 5 primary analysis 0.12 (-0.005, 0.29). Due to the uncertainty around each result, as represented by the confidence interval that crosses zero, we are unable to conclude that the class-level HHS intervention had an effect on the writing ability of Year 2 pupils or the small groups intervention had an effect on the writing ability of Year 5 pupils that struggled to write fluently.

Table 27: Primary analysis

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges g (95% CI)	p-value
WAM_CJ Year 2	1679 (196)	-0.08 (-0.21, 0.04)	1765 (214)	0.19 (0.07, 0.31)	3421 (1667; 1754)	-0.02 (-0.13, 0.10)	0.77
WAM_CJ Year 5	167 (27)	0.18 (-0.11, 0.48)	162 (24)	-0.19 (-0.47, 0.09)	313 (159; 154)	0.12 (-0.05, 0.29)	0.16

Criterion reference scores

The criterion reference scores (WAM_CR) distribution at follow-up can be found in Appendices E1 and E2 together with QQ plots of fitted versus observed residuals. The distributions of scores were summarized as histograms fitted with normal curves with the same mean and standard deviation as the underlying WAM_CR distribution. For both experiments, the distribution of the WAM_CR scores deviates from normality; the QQ plots of fitted versus observed residuals (see Appendices F1 and F2) suggest that, nevertheless, the analysis models fitted the data adequately.

The results of the analyses of the first secondary outcome (WAM_CR) are presented below, in

Table **28**. Regression models similar to the ones applied on the primary outcome analysis were applied to the writing ability data of the secondary subsample of Year 2 pupils and to the Year 5 sample. The Hedges' g effect size for the Year 2 experiment's first secondary analysis was -0.05 (-0.25, 0.16) and for the Year 5 experiment was 0.04 (-0.22, 0.213). Both confidence intervals straddle zero and as such we cannot reject the null hypotheses that the mean scores for the two randomisation groups are similar. As was the case with the primary analyses described in the previous section, the evidence does not support the case for the HHS intervention having had an effect on the writing composition of any of the cohorts assessed in the trial.

Table 28: Secondary analysis A

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges g (95% CI)	p-value
WAM_CR Year 2	230 (30)	9.67 (9.17, 10.2)	213 (42)	10.1 (9.65, 10.6)	439 (228; 211)	-0.05 (-0.25, 0.16)	0.65
WAM_CR Year 5	167 (27)	13.1 (12.6, 13.6)	162 (24)	13 (12.4, 13.5)	313 (159; 154)	-0.04 (-0.22, 0.13)	0.63

Handwriting speed test

As explained in detail in the outcome measures section above, the handwriting speed tests scores were aggregated into an ordinal variable (HST_FS) whose distributions at baseline and follow-up are described in bar charts that can be found in Appendices F1 and F2.

To evaluate the impact of the HHS intervention on the handwriting speed of Year 2 pupils and Year 5 pupils who struggle to write fluently due to poor fine motor skills, we have fitted ordinal regression models that are suited to model categorical data whose categories have an underlying implicit order. For the Year 2 experiment, we have fitted a two-level random intercepts model and for the Year 5 experiment a single level fixed effects model. The results of the analyses are displayed in Table 29 below. We fail in both cases to reject the null hypotheses that the number of full sentences a pupil can write in a three-minute window will fall in the higher or lower categories of the HST_FS variable independent of randomisation group. According to our analyses, it is unlikely that pupils who received the HHS interventions had their handwriting speed increased or decreased when compared to the pupils in the control group.

However, we do not wish to dismiss outright the possibility of any of the HHS intervention having had an effect on children's handwriting speed. Aggregating the original letter-counting HST_RS scores to sentence counting HST_FS scores lead to a loss of granulation and, consequently, information, and it may be the case that the aggregated variable is just too blunt an instrument to detect a small effect size. It should also be noted that we have not evaluated the power of the ordinal regression analyses and, as such, we cannot predict the minimum effect size that is possible to detect under these circumstances.

Table 29: Secondary analysis B

Outcome	Total n (intervention; control)	Raw intervention coefficient	Standard Error	p-value
HST_FS Year 2	437 (227; 210)	0.01	0.27	0.96
HST_FS Year 5	328 (167; 161)	-0.13	0.13	0.33

Analysis in the presence of non-compliance

As per the SAP, we have run two stages least squares instrumental variables regressions with group allocation as the instrumental variable to determine if there was an association between writing ability, as measured by WAM_CJ scores, and two of the compliance measures described below (see Compliance section).

The first measure of compliance we have analysed, Compliance 1, was a pseudo-continuous dosage measure that described for how long the children randomised into the intervention group were exposed to the intervention. Reflecting the two different settings of the intervention—class and small groups—Compliance 1 was defined for the Year 2 experiment at school level as the average total length of HHS sessions delivered per class, and for the Year 5 experiment at pupil level as the total time of exposure. The second measure, Compliance 2, was a binary measure that identified if a child, n, attended at least 20 out of 24 intervention sessions for the Year 5 experiment and if all the Year classes in an intervention school had at least 20 out of 24 class based interventions delivered for the Year 2 experiment.

The results of the two separate CACE analyses for the Year 2 experiment are summarized in Table 30 and those of the corresponding analyses for the Year 5 Experiment are displayed in Table 31. Results for the first stage of the instrumental variables regression, as well as correlations between treatment allocation and the endogenous compliance variables and their associated weak-instruments test F statistics, are reported in Appendix C1.

The CACE estimates and their associated 95% confidence intervals were converted to Hedges' g effect sizes. The confidence intervals for the four analyses ranged between negative and positive values, straddling zero. With no result could we conclude that writing attainment, as measured by comparative judgement scores (WAM_CJ), was associated with compliance.

The evidence gathered in both experiments does not support the case for an association between length of exposure to the intervention and writing ability, regardless of the mode of delivery.

For both experiments, we could not find any statistical evidence that attending at least 20 out of 24 practice sessions had an effect on the overall effectiveness of the programme. As such, we cannot corroborate or disprove the conclusions of the systematic review of Hoy et al. (2011) discussed in the protocol that suggested that interventions that included less than 20 sessions were ineffective.

Table 30: Complier Average Causal Effect (CACE) analysis—Year 2 experiment

Compliance Measure	Total n (intervention; control)	Hedges g (95% CI)	p-value
Compliance 1 (school level)	3084 (1330; 1754)	-0.003 (-0.014, 0.008)	0.59
Compliance 2 (school level)	3084 (1330; 1754)	-0.039 (-0.181, 0.103)	0.59

Table 31: Complier Average Causal Effect (CACE) analysis—Year 5 experiment

Compliance Measure	Total n (intervention; control)	Hedges g (95% CI)	p-value
Compliance 1 (pupil level)	306 (152; 154)	0.011 (-0.005, 0.027)	0.17
Compliance 2 (pupil level)	306 (152; 154)	0.144 (-0.061, 0.35)	0.17

Missing data analysis

Since we came across measurement attrition levels in excess of 5% for both the Year 2 experiment (11%) and the Year 5 experiment (5%), we ran missing data analyses.

For the purpose of investigating whether the follow-up data was missing at random (MAR) for the Year 2 experiment, a two-level (pupil and school) random intercepts logit model was fitted with a missingness flag (TRUE if WAM_CJ at follow up was missing, FALSE if present) as the dependent variables and baseline WAM_CJ scores, control/intervention indicator, region, and pupil gender as covariates. WAM_CJ was a highly significant predictor of missingness (p-value of O(10-9)), with the likelihood of missingness decreasing as the baseline score increased. We also ran a similar model considering an absence at follow-up testing flag (TRUE if pupil did not sit the follow up test, FALSE if pupil was present) as the dependent variable and the same group of covariates. Baseline WAM_CJ was once again a highly significant predictor of absence (p-value of O(10-8)), with the likelihood of absence decreasing with the baseline WAM_CJ score.

Since baseline was included in the ITT model and individual missing pupil scores are unlikely to be related to group allocation, we consider it unlikely that the Year 2 experiment analysis is biased.

We have used the same type of methodology to enquire if the follow-up data of the Year 5 experiment was missing at random (MAR). A single level logit model with a missingness flag as the dependent variable was fitted with baseline WAM_CJ scores, control/intervention indicator, region, pupils' FSM status as reported by their school, pupil gender, and school site as covariates. We have identified site as a significant predictor of missingness (p-value = 0.049). As was the case with the Year 2 experiment, the significant predictor of missingness was unrelated to randomisation group.

As described in the methods section, we have run the primary ITT analysis on a dataset with imputed baseline and follow-up missing values. The missing values were imputed using predictive mean matching, with five plausible values derived for each case. The primary ITT model was re-run on the five sets of imputed plausible values and the estimates for each model were then pooled into a single set of estimates and standard errors that was compared to the results of the original analysis. The comparison results can be seen in Table 32.

Table 32: Comparison of primary ITT model and pooled estimates from imputed values models—Year 5 experiment

	Raw randomisation group (intervention) coefficient	Standard Error
Primary ITT model	0.22	0.158
Pooled estimate	0.21	0.157

There is a good level of agreement between the results of the primary ITT model and the pooled estimates, which leads us to believe that it is unlikely that the Year 5 experiment analysis is biased due to missing values.

Subgroup analyses

As the Year 2 experiment did not require FSM eligibility as a randomisation stratifier, this data was not collected at the time. We have hence had to wait until the NPD data was available through the SRS in order to carry out the FSM subgroup analysis.

We have assessed if the Helping Handwriting Shine intervention delivered at class level might have had a differential impact on children that were FSM-eligible as compared to their counterparts by fitting two multi-level models: an interaction model and a model similar to the primary analysis model but restricted to the subsample of Year 2 pupils who were FSM-eligible in 2018 or any of the six years that predated the trial. The models are described in detail in the SAP (Andrade et al., 2019).

The results of the interaction model are displayed in Table 33, below. The raw interaction term of randomisation group with the FSM indicator had an associated p-value of 0.239. In this case, the results of the interaction model do not support the case for the class-level intervention having had an effect on the writing ability of children who were eligible for FSM in 2018 or any of their school years to the date of the trial. As specified in the SAP, we have interrogated the possibility of the small groups setting mode of delivery of the HHS intervention having had a differential effect on specific

subgroups of the Year 5 sample by fitting interaction models. We have considered the differential effect of the intervention on FSM-eligible children and on slow and effortful writers versus children whose handwriting is illegible. The choice of the latter subgroup was motivated by the eligibility criteria of the Year 5 experiment that focused on these two groups of children and we have used the baseline scores of the handwriting speed (HST_FS) as a proxy to distinguish between slow and regular/fast writers.

The description of the analysis results, as well as of the variables interacted with the randomisation variable, can be seen in Table 33 below.

The raw interaction terms of group allocation with the variable of interest had, for the two interaction models we ran, associated values that exceeded the significance level of 0.05 and interaction coefficients were low. In both cases the statistical evidence did not meet the threshold to conclude that the small groups setting mode of the intervention had an effect on FSM-eligible children or a differential effect when slow and effortful writers were compared to writers whose handwriting was illegible.

Table 33: Subgroup analysis—interaction models

Variable of interest	Raw interaction coefficient	Standard error	p-value
EVERFSM_6_P_SPR18 Year 2	0.19	0.16	0.239
Eligible for FSM (reported) Year 5	0.1	0.35	0.78
HST_FS (baseline) Year 5	−0.47 (HST_FS=1 *intervention); −1.57 (HST_FS=2 *intervention); −0.41 (HST_FS=3 *intervention); −0.63 (HST_FS=4 *intervention); −1.5 (HST_FS=5 *intervention); −0.18 (HST_FS=6 *intervention); −1.29 (HST_FS=7 *intervention)	1.43 (HST_FS=1 *intervention); 1.19 (HST_FS=2 *intervention); 1.16 (HST_FS=3 *intervention); 1.16 (HST_FS=4 *intervention); 1.18 (HST_FS=5 *intervention); 1.27 (HST_FS=6 *intervention); 1.37 (HST_FS=7 *intervention)	0.74 (HST_FS=1 *intervention); 0.19 (HST_FS=2 *intervention); 0.73 (HST_FS=3 *intervention); 0.59 (HST_FS=4 *intervention); 0.2 (HST_FS=5 *intervention); 0.89 (HST_FS=6 *intervention); 0.34 (HST_FS=7 *intervention)

A second approach, also prescribed in the analysis guidance document (EEF, 2008), was to run the primary analysis model for the subsample of FSM6-eligible Year 2 pupils. The results of the restricted model are displayed in Table 34, below. The effect size for the restricted analysis is 0.01 and its associated 95% confidence interval straddles zero. We therefore cannot reject the null hypothesis that the average pupil writing ability measured in terms of the Writing Assessment Measure's comparative judgement true scores (WAM_CJ scores) does not differ between treatment and

control groups. Once again, we found no evidence that the class-level intervention had a differential effect for children who are FSM6-eligible.

We also evaluated if the intervention had a differential impact on FSM-eligible pupils by running the fixed effects primary analysis model for the subsample of pupils reported as FSM-eligible in the Year 5 cohort. As can be seen in Table 34, this subgroup analysis is underpowered to detect effect sizes smaller than 0.44 and is, as such, merely exploratory.

The effect size for the Year 5 FSM-restricted analysis is 0.17 and its associated 95% confidence interval includes zero. We cannot reject the null hypothesis that the average WAM_CJ differs between randomisation groups. As before, we found no evidence that the intervention had a differential effect for children reported as FSM-eligible.

Table 34: Subgroup analysis—FSM-eligible pupils

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges g (95% CI)	p-value
WAM_CJ Year 2	437 (92)	0.45 (0.19, 0.70)	337 (76)	0.18 (−0.06, 0.42)	774 (437, 337)	0.01 (−0.14, 0.16)	0.868
WAM_CJ Year 5	42 (17)	−0.67 (−1.24, −0.1)	49 (7)	−0.77 (−1.36, −0.18)	91 (42; 49)	0.17 (−0.18, 0.51)	0.34

Comparison of FSM6 Spring 2018 and single-point FSM for the Year 5 Experiment

As described in the SAP, the FSM subgroup analyses for the Year 5 experiment were specified in terms of pupils' FSM eligibility status as reported by their schools and not in terms of everFSM or FSM6 as recommended in the EEF's analysis guidance document (EEF, 2018). This choice of design was dictated by the necessity to avoid collinearity in the analysis regression models since everFSM and FSM6 are known to be highly correlated with the reported FSM eligibility indicator. As specified in the SAP, the latter had to be included as a covariate in the Year 5 models since it was a randomisation stratifier. To assess the implications of this the evaluation team has suggested the inclusion of a table describing the overlap of the different measures.

The overlap between the trial's reported FSM indicator and the National Pupil Database variable EVERFSM_6_P_SPR18 is described in Table 35 below. The degree of accordance between the two variables is 85.1 per cent[31]. Given the missing data in the NPD field, the degree of accordance may in fact be higher. We are confident that using the reported FSM indicator as compared to FSM6 did not influence our conclusions from the Year 5 experiment.

Table 35: Overlap between Reported FSM for the Year 5 Experiment and EVERFSM_6_P_SPR18

	EVERFSM_6_P_SPR18		
Reported FSM Eligibility	Yes	No	Unknown
Yes	101	*	*
No	*	214	*

*Supressed cells, their combined figures add up to 55.

[31] Agreement in 315 out of 370 cases.

Estimation of effect sizes

The description of the effect sizes associated with the different analyses included in this report and their derivation can be found in table C2.1 (the Year 2 experiment) and C2.2 (Year 5 experiment) in Appendix C2.

Implementation and process evaluation results

This process evaluation draws on qualitative data from four training observations, four programme delivery observations, interviews with 24 staff (12 face-to-face and 12 via telephone) and 36 pupils from 12 schools, and interviews with three of the University of Leeds development team. It is also draws on quantitative Helping Handwriting Shine training attendance register data and quantitative and qualitative data collected from all schools via a baseline proforma, activity logs from intervention schools and business as usual logs for Year 2 control classes and Year 5 target control pupils. In the following sections, we report on implementation compliance, fidelity, and usual practice.

Compliance

Helping Handwriting Shine is a structured programme that school staff are trained to implement. The programme lasts eight weeks and consists of three 30-minute sessions per week (24 sessions in total). Following this period, HHS should be embedded into weekly teaching practice.

In this section we report on implementation compliance—whether the intervention was delivered as intended in all intervention schools. We explore the three specific compliance measures (as set out in the SAP), any issues with compliance, and why they may have occurred. We also present additional IPE data which relates to these aspects of compliance to provide further context.

The compliance model was designed jointly by NFER and the developer and consists of the following three measures:

- **attendance at training**—this measure was included because it gives an indication of training dosage for school staff;
- **number and length of handwriting sessions delivered**[32]—included because it gives an indication of how well the schools complied with the prescribed modes of delivery of the programme; and
- **extent to which schools use the programme after the eight-week delivery**—included because it gives an indication of reach and responsiveness.

Attendance at training

A minimum of two members of staff from each school delivering to Year 2 and Year 5 intervention classes/groups were required to attend a one-day training session provided by the developer prior to delivery. At least one member of staff attending from each school was required to be a senior leader to be fully compliant to the training requirements of the intervention (as per the SAP).[33] As cascading is not a recognised feature of the intervention, schools were encouraged, where possible, to send all those nominated to deliver the intervention to the scheduled training.

The training commenced in the third week of October 2018. Eight training sessions were delivered in total: two session dates were scheduled in each of the four regional hub locations (Leeds, Sheffield, Darlington, and Newcastle).[34] A small number of staff (seven) from three schools were unable to attend one of the scheduled sessions and the developer agreed to provide additional training sessions—face-to-face and by Skype—for those staff on three further occasions (around two weeks after the final scheduled training event, towards the end of November 2018). A record of attendance at each training session was collected by the developer; these were collated and passed to the research team for analysis.

Overall, 46 intervention schools—100% of schools not withdrawn at this point—attended the training and 146 members of staff were trained in total.[35] On average, schools sent three members of staff to the training (the maximum number of staff sent by schools was six and the fewest was two); 43 (of the 46) schools that attended met minimum requirements

[32] This measure also gives an indication of intervention dosage for pupils. This forms the basis of the CACE analysis and is reported fully in the section on Analysis in the presence of non-compliance.
[33] The rationale for this was so that senior leaders fully understood the nature and requirements of the intervention.
[34] NFER researchers observed half of the training sessions and found the delivery to be thorough (see Quality of Training section below). There were good levels of school engagement and responsiveness across the sessions observed.
[35] There were 139 attendees at the hub training events and seven at alternative sessions.

to send at least one staff member responsible for delivering to Year 2 and Year 5 (for Year 2 this was typically the class teacher(s) and for Year 5 this was typically TAs).[36] Most of the schools attending the training sent a member of their SLT (such as the headteacher or a deputy/assistant headteacher) or another senior member of staff (for example, the KS1 lead or SENCO) and were categorised as having 'high compliance' to requirements (as per the SAP). Four schools did not have any senior leader representation and these schools were categorised as having 'low compliance' (but are still included within the impact analysis). Six staff members from different schools registered to attend the training (that is, their names were included on the attendance log) but they were recorded as not present; these staff were retained on the developers contact list. Of these, three were headteachers who were unlikely to be directly involved in delivery of the intervention, two were TAs, and one was a Year 2 teacher.

There was evidence from the IPE that, in a small number of cases, staff who were unable to attend the training went on to deliver HHS in intervention schools. This included a very small number of schools that asked to cascade to other members of staff (as observed by the research team at training events). These particular schools were keen to commence delivery as soon as possible in order to complete a four-week block of intervention delivery before the half-term school holiday. It was decided jointly between the developer and NFER that while this was not to be encouraged more widely, it would be permitted in the interim while follow-up training was being arranged for these schools. Interviewed members of the University of Leeds team also reported that they were aware, through their ongoing contact with schools, that, in a small number of cases, trained staff members had left during the intervention period. Training for replacement staff was offered but this was not taken up. In addition, while all of the school staff interviewed had attended the training, one interviewee reported that an untrained member of staff in their school was delivering the intervention to pupils in Year 5.

Overall, these findings suggest that in the majority of schools, trained staff delivered the intervention, providing good evidence for the theory of change that the determined conditions of training were mostly adhered to.

Quality of training and further support

In addition to this specific measure of compliance, the IPE explored the extent to which attendance at the training prepared school staff for delivering and managing HHS. All of the case-study interviewees were positive about the training they received and believed it had equipped them well to deliver the intervention in reality—finding it clear and comprehensive. Role-play activities, which involved, for example, attendees delivering a HHS session to colleagues, were considered particularly helpful. These activities enabled participants to practice delivery as well as gain an understanding of the recipient's experience (reported by a third of case-study interviewees). The fact that the training closely followed the intervention manual helped attendees to fully understand each stage of the intervention and what was expected of them (reported by around a quarter of case-study interviewees). Other elements of the training that facilitated its effectiveness (reported by a small number of interviewees) were:

- knowledgeable trainers who explained the intervention well and who were able to answer attendees' questions;
- the opportunity to gain an understanding of the theory relating to the pedagogy of handwriting;
  the appropriateness and accessibility of the training for both teachers and support staff; and
- the opportunity to discuss with/hear from staff from other schools.

The IPE also investigated the extent to which schools accessed other ongoing training and support from the developer to facilitate their delivery of the intervention. This support involved follow-up, half-day CPD embedding training in each hub location,[37] contact with the development team (via telephone, Skype, and email), and a website with access to digital materials, instructional videos, and a troubleshooting forum. Around half of the case-study interviewees stated that they had not accessed any further support after the initial training day acknowledging this was because they found

[36] The role of staff delivering the intervention (and attending the training) was not specified as a compliance measure to schools, although in Year 5, to enable the delivery of the intervention to small groups and to avoid contamination to Year 5 control pupils, it was recommended by NFER that schools utilised non-teaching staff.
[37] As attendance at the follow-up CPD was not a pre-specified compliance measure there is not enough data to draw any conclusions on attendance or responsiveness at these events more broadly. An NFER researcher observed one follow-up CPD session.

it (together with the intervention manual) sufficient.[38] Where support was accessed, this was usually in the form of regular emails distributed by the developer to all schools. To a lesser extent, interviewees reported proactively seeking out information and accessing resources via the HHS website. Individual case-study interviewees suggested some additional support that they would have found helpful; this included lesson observations and feedback on their delivery within the classroom (both of these were actually made available to schools but not accessed), further information and strategies for how to effectively deliver at a whole-class level to Year 2 pupils with a range of ability levels, and advice on marking/assessing pupils work (and their progress) as part of the intervention. Overall, this suggests that ongoing support from the University of Leeds team was sufficient to support most school staff to deliver and embed the programme. In addition, analysis of the small number of completed monthly post-programme logs[39] showed that no intervention schools reported receiving other external support or CPD for handwriting specifically, or for literacy more generally, from the end of the eight-week programme until the pupil endpoint testing.

Number of sessions delivered

Analysis of interviews with case-study schools reflected a similar pattern to the one described for Compliance Measure 2 in the 'Analysis in the presence of non-compliance' section. Around three quarters of staff (Year 2 and Year 5) reported that they always delivered three sessions per week. Other schools were not always able to manage three sessions each week despite their best efforts. Missed sessions were due to time constraints within an already busy curriculum timetable, staff availability and capacity issues, requirements for pupils to participate in other school activities (such as one-off events or plays), and staff and pupil absence. These same issues and challenges were identified by schools in the post-intervention reflections log.

These findings provide good evidence for the theory of change that this predetermined condition of dosage was adhered to. The dosage across the sample is fairly consistent although a minority of school staff did not deliver the intended three sessions per week, which threatens the intervention fidelity.

In addition to this specific compliance measure on session dosage, the IPE also investigated if schools adhered to trial requirements to deliver the required number of sessions to the whole class for Year 2 and in small groups for Year 5. All of the case-study interviewees delivering to Year 2s stated that their sessions were delivered to the whole class.[40] There were some rare instances when individual Year 2 pupils with high-level SEN did not participate but remained in the classroom engaged in another activity. The way Year 2 classrooms were organised for the delivery of HHS sessions varied across schools. This included, for example, pupils seated together in pre-specified groups based on ability levels as well as pupils assembled in specific literacy groupings (for handwriting, guided reading, or other activities). All but one of the Year 5 case-study interviewees stated that they delivered all sessions to Year 5 intervention pupils in small groups outside of the Year 5 classroom (as advised by NFER in order to minimise contamination to target control pupils within the class).[41] Year 5 sessions typically took place in a separate teaching room dedicated to intervention delivery or in another learning space such as the school library. In other case-study schools, however, such spaces were limited;[42] this included one school where pupils received HHS in a corridor and another school where the sessions were delivered in the staffroom (as observed by a researcher). The limitations of delivering the sessions in such spaces is that they may not be fully conducive for handwriting (for example, where pupils require appropriate desks and seating). In the case-study school where delivery took place inside the Year 5 classroom, the Year 5 teacher delivered the intervention due to the limited availability of support staff in the school. Aware of the need to separate intervention and control pupils during delivery, this teacher was conflicted by the desire to involve all pupils who needed handwriting support while at the same time adhering to the requirements of the trial:

> *I've got a nice big classroom and I've got a table set up behind some bookshelves and we sit over there while the rest of [the class] are doing accelerated reading … The rest of the class have picked up a bit [of the intervention] because I do it in the corner. I hear them say, 'That's my favourite warm up.' So they have picked*

[38] Three of the nine schools that completed Year 2 post-programme logs had accessed further support from the developer and two of the nine schools that completed Year 5 post-programme logs had accessed further support from the developer.
[39] Nine Year 2 class logs and nine Year 5 group logs: the nine responding schools in the two groups were different.
[40] The Year 2 sessions observed by researchers were all delivered this way.
[41] The Year 5 sessions observed by researchers were all delivered outside of the main classroom. Clear guidance about avoidance of contamination was given to schools in written communication from NFER (this included a 'Dos and Don'ts' poster to display in school) and verbally by the developer at the initial training day.
[42] This was also identified as challenge by a small number of Year 5 staff in the post-intervention reflections log.

up on what's going on. There may have been contamination. To be honest, I was desperate to contaminate the whole class. I can see that this is something that is really going to benefit the whole class.

The developers also described similar contamination issues during interviews with the research team. Schools had made them aware of Year 5 sessions being delivered in the main classroom as well as peer-to-peer sharing which occurred independently between Year 5 pupils. While none of the schools that completed the Year 5 monthly post-programme logs recorded that control pupils had heard about, or saw, any HHS work or teaching, the overall response rate was low. It is therefore difficult to determine to what extent contamination was widespread.

An issue that arose during the 'live' trial was delivery time. It was intended—and intervention schools were informed—that half of the sessions must be delivered before the Christmas break and half afterwards, but this was difficult for some schools: they reported finding it hard to fit the sessions into the few weeks before the end of the term due to additional school commitments such as shows, trips, and so forth. On receiving this information, we encouraged schools to restart the intervention as quickly as possible, but also gave them the flexibility to deliver whatever remained of the intervention as opposed to only delivering what they could within the specified time period. This meant that all schools completed the intervention but had the knock-on impact that some schools finished early in February, and some not until late March.

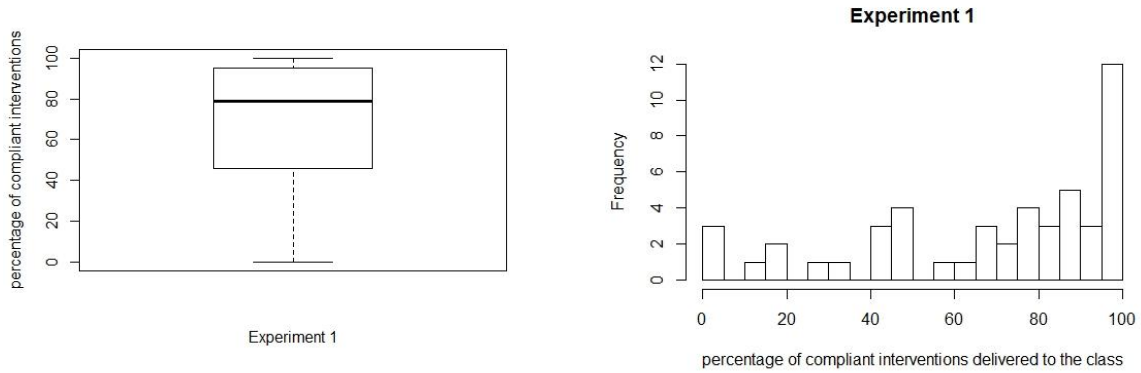
The implications of this could, speculatively, have included differences in the amount of engagement from schools in the transition period and their capacity to fully invest in a transition action plan. It may have resulted in some lack of attendance at the half-day training; it may also have resulted in further deviation in the type and extent of embedding during the gap period.

Session length

In order to be compliant with the HHS programme, pupils should receive three 30-minute sessions each week for the duration of the intervention period. The delivery log asked Year 2 and Year 5 staff to record the length of each session they delivered to their respective pupils over the eight-week period. To aid completion, they were provided with the categories 'not delivered', 'less than 30 minutes', '30 minutes', and 'more than 30 minutes'.

We have analysed the data collected in 49 classes of the 43 intervention schools that submitted intervention logs: 15% of the total number of class-level sessions delivered to Year 2 lasted less than 30 minutes, 16% lasted more than 30 minutes, and 69% complied with the 30 minutes duration. We have also measured, for each Year 2 class, the percentage of the total number of sessions delivered that complied with the prescribed 30-minute duration. The results are summarised in a boxplot and in a histogram below (Figure 8). On average, each intervention Year 2 class complied with the 30-minute delivery in 69% of the sessions delivered. Furthermore, approximately seven out of ten classes complied with the 30-minute delivery in at least 50% of the sessions delivered and approximately half of the classes complied with the prescribed delivery time in approximately 80% of the sessions delivered. A total of eight Year 2 classes, approximately 16%, were fully compliant and kept to the 30-minute time frame in all the sessions.

Figure 8: Boxplot and histogram for the distribution of percentage of compliant (30-minute) sessions delivered to Year 2 classes

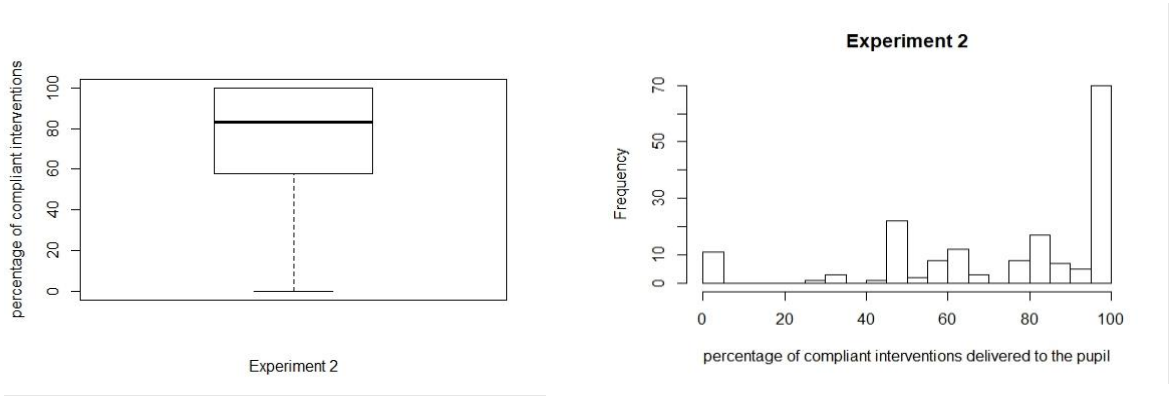


For the Year 5 small-group intervention, we have analysed compliance data for 170 children who took part in the trial and whose attendance information was collected in intervention logs. Although two of these children were randomised

into the control group and one intervention pupil stopped attending the school during the trial, their information was nonetheless considered suitable to be included in this analysis. These children did attend intervention sessions and the information indicates how well schools were complying.

Thirteen percent of the total number of pupil-level intervention sessions delivered to Year 5 pupils lasted less than 30 minutes, 13% lasted more than 30 minutes, and 74% were compliant at 30 minutes. Similarly to the Year 2 experiment, for each pupil we have also calculated the percentage of the total number of individual intervention sessions that were compliant with the pre-specified 30-minute delivery. The results are displayed in a boxplot and in a histogram below (Figure 9). On average, Year 5 children received compliant sessions in 76% of the sessions they attended; 77% of the Year 5 children attended 30-minute sessions at least 50% of the time, and a little over 40% of the children (70 out of 170) only attended compliant intervention sessions.

Figure 9: Boxplot and histogram for the distribution of percentage of compliant (30-minute) sessions delivered to Year 5 pupils



Our analysis suggests that the levels of compliance in terms of session length were high for both experiments, with those of the Year 5 experiment exceeding those of the Year 2.

Around three quarters of case-study staff interviewed stated that their HHS sessions typically lasted 30 minutes. Those who reported shorter sessions stated that this was largely due to the needs and characteristics of pupils in their classes/groups, which meant they were not always able to hold their attention fully for the allotted time.[43] Despite this, they did not appear to deviate significantly from the suggested 30-minute timings. One case-study interviewee delivering to Year 5 pupils, for example, described how their sessions usually lasted slightly longer in order to allow pupils to settle after transitioning from their usual classroom to the intervention delivery space. The interviewees who did not always comply with the 30-minute session duration were mostly Year 5 staff, half of these interviewees also reported non-compliance with the number of sessions delivered each week (as reported above).

This suggests that there is only some deviation from the 30-minute time limit providing good evidence for the theory of change that this pre-determined condition of dosage was mostly adhered to.

Support provided to pupils after the intervention period

The developers supported schools in embedding HHS into weekly teaching practice (through the use of gap tasks and extension tasks), helping to ensure that intervention pupils continued to receive frequent support 'little and often' after the formal eight-week intervention period. The evaluator's initial intention was to track schools' integration of the intervention after the formal delivery period until the pupil endpoint testing via a monthly post-programme log. However, as previously described in the methodology section, we did not actively pursue the completion of this log in order to reduce the burden on schools. The small number of completed post-intervention logs that we received (nine Year 2 logs and nine Year 5 logs)[44] showed that while most of the responding schools regularly used HHS activities, techniques,

[43] This was also identified as a challenge by a small number of schools in the post-intervention reflections log.
[44] The nine responding schools in the two groups were different.

and gaps tasks in the first month after the formal intervention period, this diminished over time. Three Year 2 schools and two Year 5 schools used the techniques ‘occasionally’ after March 2019.

Staff interviewed during case-study visits, which took place during the intervention period (and therefore before the half-day training events that were provided to support schools in the transition), were asked about their plans for embedding. Year 2 staff (mostly class teachers) were generally committed to the prospect of embedding, concluding that it would be difficult to continue to deliver the intervention to the same intensity for a longer period. None of the Year 2 interviewees had particularly clear plans for embedding at this stage, but they expressed a desire to continue to implement particular features of the intervention that they felt had enhanced their usual teaching practice (for example, the preparing for writing activity and warm-up tasks). Generally, case-study interviewees who were delivering to Year 5 pupils (mostly TAs) were unclear about embedding arrangements at that time, acknowledging that this would typically be a future decision for teaching staff or the SLT. Unfortunately, the timing of the case-study visits did not enable us to gather information on how these schools approached embedding after the CPD events (nor whether they attended them).

Telephone interviews with staff in half of the case-study schools took place in June and July (around the time of the pupil endpoint testing, four or five months after delivery of the eight-week programme had been completed) and staff were asked if, and how, they had continued to embed HHS. Embedding activities had taken place in most Year 2 classes and involved the delivery of warm-up activities and gap tasks. One Year 2 teacher continued to deliver one 30-minute session per week to the whole class and another delivered a weekly small group session to Year 2 pupils after school (which was perceived to be more manageable). Two Year 2 teachers acknowledged, however, that embedding activities had decreased over time as preparation for KS1 national curriculum tests took priority. Other than the occasional verbal reminders to Year 5 pupils about their writing in class, the Year 5 staff interviewed had not continued to embed any aspects of HHS. However this, in part, is likely to be a consequence of the within-school trial design and the requirement to avoid contamination of target control pupils (as discussed previously).

The developers also assert that, as part of the intervention, senior leaders should host weekly ten-minute progress meetings with delivery staff in Year 2 and Year 5 during the eight-week delivery period—one of the reasons for this was to help ensure the intervention was prioritised. Just over half of case-study interviewees delivering HHS in Year 2 and Year 5 met periodically to briefly discuss delivery and pupil progress. In some cases, the Year 2 teachers were also senior leaders or literacy coordinators. There was no evidence from the case studies, however, that senior leaders who attended the training but who were not directly involved in delivery of the intervention were involved in weekly progress meetings; this may have had implications for the embedding of the programme in the longer term in these schools.

Data suggests some variation in the extent and duration of ongoing embedding activities across schools and particularly in Year 5. There is not strong evidence for the theory of change that this pre-determined condition of delivery was adhered to by most schools. This area is likely to require further research.

Altogether, the findings on the three specific compliance measures indicate that not all schools ran HHS exactly as intended, particularly the delivery of the required number of sessions and the ongoing embedding of activities. The evaluation does, however, appear to capture the real-life running of the programme.

Fidelity

Adherence to the HHS structure was consistently stressed in both the training and the intervention manual. The developers identified five key principles, aspects of the intervention that were specified as non-optional for schools:

- frequency of practice—HHS sessions should be delivered three times per week;
- structure—each HHS session should comprise three parts: preparing for handwriting (5 minutes), a warm-up activity (5–10 minutes) and a handwriting activity (15–20 minutes);
- variability of practice—handwriting should be practiced in a variety of ways to enable generalisability;
- metacognition—each activity should involve a ‘model-plan-evaluate’ structure (that is, a 'thinking about thinking' approach) to facilitate the acquisition of skills; and
- engagement with support.

We report on schools' implementation across these elements in the following sections. We also present additional IPE data on permitted elements of adaptation.

Adherence to the structure of the intervention

Preparing for handwriting is the first part of the intervention; its purpose is to get pupils physically and mentally ready for writing. It involves gathering pupils' attention and addresses their seated position, posture, paper angle, pencil grip, and pressure. The developer specifies that preparing for handwriting should be delivered at the start of every HHS session. In the delivery logs, completed during the eight-week intervention period, schools were asked to record if they allotted five minutes to preparing for handwriting during each session delivered (at the class level for Year 2 and at the pupil level for Year 5). This element of the intervention was reported as mostly delivered as intended to both groups. Just under a quarter of Year 2 staff did not always adhere to this element of the intervention. The allotted five minutes for preparing for handwriting was not delivered to six Year 2 classes on one or two occasions over the eight weeks, which probably reflects the inevitability that, on occasion, staff will miss or alter delivery because of unexpected issues in class. In five Year 2 classes, however, this element was not delivered as intended on a more regular basis (between six and 18 of sessions delivered). We found a similar pattern for Year 5 delivery. Overall, 17% of Year 5 intervention pupils were not given this time for preparation in at least one of the sessions they received. We should treat these findings with caution, however, as they do not necessarily mean that schools missed out this element entirely. Indeed, some case-study interviewees report that the preparing for writing element of the intervention became faster as pupils grew more familiar with the process. In such cases, additional time was allocated to warm-up and handwriting activities (rather than the overall session length being reduced).

The second part of a HHS session is the *warm-up activity*. This is a short pen or pencil activity that involves practising the types of skills pupils need for writing. The developer specifies that one warm-up activity should be delivered during each session. In the delivery logs, schools were asked to record if they delivered a warm-up activity during each session of the eight-week intervention period (at the class level for Year 2 and at the pupil level for Year 5). Almost all staff reported delivering HHS to Year 2 classes provided a warm-up activity in every session; just two Year 2 classes did not. One class missed the warm-up activity twice over the eight weeks and the other class missed a warm-up activity six times (analysis shows that this Year 2 teacher was also regularly non-compliant with the preparing for handwriting activity). All Year 5 intervention pupils were reported to have participated in a warm-up activity during every session delivered. These findings suggest that this element of the intervention was delivered as intended to Year 5 pupils and mostly as intended to Year 2 classes.

The third and final element of a HHS session is the *handwriting activity*—the main focus of the session. It involves a longer task where pupils directly practice handwriting skills. The developer specifies that one writing activity should be delivered during each session. In the delivery logs, schools were asked to record if they delivered a writing activity during each session of the eight-week intervention period (at the class level for Year 2 and at the pupil level for Year 5). Almost all staff delivering HHS to Year 2 classes provided a writing activity in every session; just one class missed the handwriting activity on one occasion. Our analysis shows that the Year 2 staff member responsible for delivery in this school largely adhered to the specified structure of the intervention, only missing two warm-up activities and not delivering the preparing for handwriting element as intended once during the eight-week period. All Year 5 intervention pupils participated in a writing activity during every session delivered. These findings suggest that this element of the intervention was delivered as intended to Year 5 pupils and mostly as intended to Year 2 classes.

All of the case-study interviewees reported delivering the three-part HHS structure following the clear instructions to do so from the developers in the training and as set out in the intervention manual.

These findings suggest that the majority of schools kept to the structure of the intervention, providing good evidence for the theory of change that this pre-determined condition of delivery is adhered to by most schools. There are, however, some potential limitations with the delivery logs, (discussed further in the limitations and lessons learned section), which mean that these findings should be viewed with caution. School staff were encouraged to complete the logs after each session. However, we were made aware, anecdotally, that some completed them retrospectively on a weekly, or longer, basis. This may have affected their ability to recall small variations in delivery on a session-by-session (for Year 2) or pupil-by-pupil (for Year 5) basis. A further limitation of the delivery log for Year 5 pupils was identified at the analysis stage. In most cases, there appeared to be little variation in the individual logs of pupils within the same Year 5 group. While this could reflect actual delivery, it could indicate that in order to save time on completion, staff entered the same information for each pupil, rather than recording on a truly individual basis.

Variability of practice

HHS requires pupils to practice handwriting tasks in a wide variety of ways. Included in the intervention manual are six warm-up and five handwriting activities. The developer specified that no activity should be repeated in any given week of delivery to ensure sufficient variation. The intervention manual provides an example curriculum to support this process. In the delivery logs, schools were asked to record which warm-up activity (1–6) and which writing activity (1–5) was delivered during each session (at the class level for Year 2 and at the pupil level for Year 5).

Findings from the logs suggest there is good evidence for the theory of change and the implementation of this pre-determined condition of delivery. Most intervention pupils received the required range of warm-up and writing activities over the eight-week intervention period, although a small subgroup of schools did not adhere to requirements to deliver a variety of tasks each week:

- All Year 2 classes and Year 5 pupils were given a variety of *warm-up activities* (1–6). One in every ten Year 2 classes and around one third of Year 5 pupils, however, experienced repetition of a warm-up activity within the same week (on at least one occasion).
- Almost all Year 2 classes and Year 5 pupils received a variety of *writing activities* (1–5).[45] A third of Year 2 classes and just over a quarter of Year 5 pupils, however, received the same writing activity more than once in the same week (on at least one occasion).

All case-study staff interviewed reported that they varied the warm-up and writing activities they delivered over the eight-week intervention period. Indeed, several explained that they chose to adhere to the example curriculum in the manual as this provided an easy way to ensure they followed an established pattern. Some interviewees did, however, comment that there were certain activities that pupils enjoyed, such as the shading and spiral warm-up tasks, which they used more than others.

Use of the 'model-plan-evaluate' cycle

HHS involves a 'model-plan-evaluate' cycle. This verbal, self-reflective technique is intended to instil metacognitive problem-solving strategies in children. The developer specifies that staff use this method with all activities; a model-plan-evaluate table was provided for each activity. In the delivery logs, schools were asked to record if they used the model-plan-evaluate process in each session. Just over three quarters of Year 2 staff always implemented it. Similarly, just over three quarters of Year 5 pupils were instructed to follow this cycle during every session they received. In the Year 2 classes where the metacognitive element of the intervention was not always carried out by staff, this occurred in between one and nine of the 24 sessions they delivered during the eight-week period. Further analysis shows that the Year 2 staff who occasionally missed this element of the intervention were generally very compliant with other aspects of the intervention.

The metacognitive element of the intervention was not always carried out by staff in case-study observations undertaken by members of the research team. One explanation, provided by case-study interviewees who initially implemented the model-plan-evaluate cycle as instructed but adapted this process over time, was that pupils became familiar with the routine and it became embedded in their own practice. Describing how the process evolved over the eight-week intervention period, a Year 2 teacher said:

> *It [the model-plan-evaluate process] was helpful in the beginning, when [the pupils] were learning and being introduced to each activity. I think as we went on, I didn't need to do that as much and as clearly, because [the pupils] were used to doing the intervention and knew what they needed to do next.*

The model-plan-evaluate process is a key component of the theory of change for this intervention. Not implementing this element as intended in each session presents a significant risk to fidelity and may mean that pupils do not fully automate the process as required. These findings provide some evidence for the assumptions underpinning the theory

[45] Two Year 2 classes and one Year 5 pupil did not; we have no further information as to why.

of change: schools mostly adhere to the process and indicate that it can become embedded in pupil practice within a relatively short time period, though evidence is limited.

Extent of adaptations that threaten intervention fidelity

Following the eight-week intervention period, all Year 2 and Year 5 staff were invited to complete a post-intervention reflections log. They were required to record how frequently they made adaptations to the programme (such as changing the content, order of activities, or purposefully missing anything). Frequently adapting in this way presents a significant risk to fidelity. Most of those Year 2 staff who responded (around three fifths) said they made adaptations occasionally. However, around one in every ten frequently adapted their sessions. Just over half of Year 5 staff said they occasionally made adaptations to the programme, although around one in every ten frequently made adaptations. The main reason for adaptations (for both groups) was to better suit pupils' needs and ability levels.

Permitted elements of adaptation

The developer identified five adaptable aspects, specifying to schools in the training and in the programme manual where permitted adaptation of the HHS intervention can be made. The following aspects were allowed to be approached with flexibility:

- Handwriting scheme and styles—schools could adapt the intervention to fit whatever scheme of handwriting or style they usually followed.
- Levels—each warm-up and writing activity is divided into three levels, with level one being the easiest and level three being the hardest. Schools are able to use their judgement as to which level is most suitable. It was recommended that each pupil receive the intervention at a level appropriate for them. It was okay if a pupil never passed a certain level as long as it was an appropriate level of challenge.
- Extension tasks—schools were encouraged to extend activities for those pupils who progressed beyond level three tasks if they deemed it appropriate for the child. An example extension task was provided for each activity and session leaders could make up their own.
- Gap tasks—schools were encouraged to administer these in between formal intervention sessions in order to embed practice in other lessons and at home. The frequency and level (difficulty) of the gap task pupils must complete could differ at the teacher's discretion.
- Rewards—pupils or group rewards could be adapted to fit with the school policy.

The following findings draw on data recorded by schools in delivery logs, completed each week during the eight-week intervention period (at the class level for Year 2 and at the pupil level for Year 5), and school staff interview data.

Permitted tailoring to individual levels did not feature strongly in either Year 2 or Year 5 groups. Very few Year 2 staff used tailored levels in every session—this only occurred in around one fifth of Year 2 classes, and consisted of the staff tailoring the level of the task to the individual student as opposed to delivering the same level to the whole group. Generally, Year 5 pupils received the same level tasks as others within their group,[46] although the level of tasks did usually vary (for the whole group) over the course of the eight-week intervention period. Around a quarter of case-study interviewees explained that the reason they did not tailor levels was that they found it difficult to manage, particularly at the whole-class level. Describing their approach to using levels, a Year 2 case-study teacher who was interviewed said:

> *We did it as a whole class because it was too hard to differentiate between groups. When we started delivering, we started at level 1, then when we did that task next time, we'd do the next level and build on each session, but it was a whole class thing.*

Permitted tailoring of extension activities was more common in small groups in Year 5 than at the whole-class level in Year 2. Just over two fifths of Year 2 classes received between one and nine extension tasks during the eight-week period compared to nearly half of Year 5 pupils. This could indicate that the programme was more individualised for

[46] This finding could reflect a limitation of the Year 5 logs (as reported in the section above).

Year 5 pupils as the result of the small group delivery and the increased capacity for one-to-one support, which enabled them to make faster progress. Indeed, a small number of case-study interviewees explained that the reason they did not extend activities for pupils who progressed beyond level three tasks was due to time limitations (note that extension tasks and higher levels were not mandatory).

Gap tasks were delivered to Year 2 classes by around two thirds of schools (at least once) during the eight-week intervention period. The number of gap tasks delivered to Year 2 classes ranged from 1–27. In interviews with case-study school staff, some Year 2 teachers reported delivering gap tasks occasionally, while others were optimistic that more gap tasks would be delivered as they began to embed the intervention. There was not an expectation that Year 5 staff would deliver gap tasks given the within-school trial design and the need to avoid contamination to target control pupils, though it was suggested in the training and the manual that gap tasks could be set as homework for Year 5s. Year 5 staff were, though, not asked to provide data on gap tasks in the delivery logs.

Together, these findings indicate that some schools made permitted adaptations, particularly through the use of extension and gap tasks although the extent to which they did this was variable.

Pupil engagement with intervention

In the delivery logs, Year 2 staff were asked to record class engagement, pre-specified as either 'high', 'medium', or 'low' for each session delivered. Generally, there were good levels of engagement reported among Year 2 pupils—around two-thirds of schools never logged low levels of class engagement in any of the sessions they delivered. Around a third of schools reported low levels of engagement in between one and six of the class sessions they delivered over the eight-week period. Analysis showed that some of the 'low engagement' sessions occurred towards the end of the first four-week block of delivery in the lead up to the Christmas holiday in December 2018. Staff delivering the HHS intervention in small groups to Year 5 pupils were also asked to record engagement in the same way ('high', 'medium', or 'low') but at an individual pupil level. Again, levels of engagement were generally good. High engagement was reported for almost three fifths of pupils; medium for almost a third, and low engagement was reported for one in every ten Year 5 pupils.

In our interviews with Year 2 and Year 5 pupils they mostly 'liked', or 'didn't mind' participating in HHS sessions. Elements of the programme that were considered particularly engaging by some of the staff and pupils we interviewed were the 'fun', 'enjoyable', and 'calming' warm-up activities, the competitive element of activities (that is, seeing how much of a task could be completed within a given time between peers), and the metacognitive process where pupils evaluated their own work. A small number of case-study interviewees (staff and pupils) provided reasons for lack of engagement where this occurred. This included:

- the repetitiveness of the processes and of particular activities, which meant that over time some pupils became bored and easily distracted;[47]
- experiencing fatigue or hand pain during the 30-minute session;
- the perception that some of the warm-up and writing activities were too easy and better suited to younger pupils (both Year 2 and Year 5);
- warm-up and writing activity levels were not always tailored and therefore not sufficiently differentiated to enable some pupils to engage appropriately; and
- pupil randomisation to the Year 5 group did not lead to effective group dynamics.

School readiness and foundations for the intervention

Case-study interviews with Year 2 and Year 5 school staff provided insight into the environments in which the interventions were delivered. Almost all of the case-study teachers interviewed stated that improving handwriting was a key priority for their school and that this was the main reason their school wanted to be involved in the trial. The desire

[47] This was also identified by schools as a challenge by a small number of schools in the post-intervention reflections log.

to adopt a more consistent approach to teaching handwriting across the school through the implementation of a more structured and formal process for teaching handwriting was also mentioned by a small number of the case-study teachers with whom we spoke. Interviewed teachers in two case-study schools were also explicit about the need for their schools to employ a specific intervention to improve pupils' handwriting speed and fluency and fine motor skills. Individual interviewees also reported that their schools became involved with HHS as they wanted to adopt a more interesting way of teaching handwriting—and as they had had previous involvement in EEF trials.

The HHS intervention was a good fit with case-study schools' wider strategies and plans for improvement. In almost all cases, the teachers we interviewed reported that handwriting was a key priority for their school and featured in their school improvement plan (SIP), for example. For some, there was also the intention that if proved effective, HHS would be adopted across the school in the longer term. Interviewed teachers stated that they had the capacity and readiness to take on HHS although some had initial concerns that fitting three 30-minute sessions per week into the current timetable would be a challenge. However, it was generally felt that as this was a time-limited intervention this was manageable. A lack of support staff was reported as a further limitation, particularly for the Year 5 intervention, by some interviewees in response to other areas of questioning as reported above.

Usual practice

In order to understand schools' usual practice prior to randomisation, we asked all schools to complete a baseline proforma in June 2018. They were required to outline basic details of additional literacy support for Year 2 and Year 5 classes beyond their normal curriculum and scheduled lessons. We received 103 completed proformas. Following randomisation, schools in the Year 2 control group and intervention schools with Year 5 control pupils were asked to provide information in a business as usual log. They recorded details of any general literacy or specific handwriting interventions and any handwriting practice sessions provided to Year 2 control classes and Year 5 target control pupils during the period 1 November 2018 to 31 January 2019 (this timeframe was chosen in order to align with the intervention delivery period). We received 46 Year 2 logs and 35 Year 5 control group logs. The case-study interviews also asked about usual practice in intervention schools.

In this section, we report on usual practice for Year 2 classes in intervention schools, for Year 2 classes in control schools, and for Year 5 intervention and control group pupils (in intervention schools). We describe the changes between planned support and what was delivered and the elements of HHS that differentiate it from usual practice.

Usual practice for Year 2 classes in intervention schools

The baseline proforma showed that just over three fifths of Year 2 intervention schools intended to provide some form of additional literacy support to Year 2 pupils.

- The main form of additional literacy support planned for Year 2 intervention pupils involved **specific literacy programmes** (for example, Read Write inc. and Lexia).
- The second most frequently cited planned additional literacy support was **general literacy support** (involving non-commercial or non-branded interventions) such as 'booster classes' and additional time with a TA to focus on individual literacy needs.
- The third most frequently cited form of planned support was the use of specific **reading interventions or resources** (for example, Reading Recovery and Toe by Toe).
- Just four intervention schools said they would be delivering specific **handwriting interventions** (for example, Letter Join and Rapid Writing).

Just over one fifth of intervention schools responding to the baseline proforma planned to use a mix of additional literacy support systems with their Year 2 pupils. In the majority of cases, this planned support was to be targeted at specific Year 2 pupils identified as requiring additional support and provided by a combination of teachers and TAs. One of the main differences between the intervention and usual practice in these schools was that HHS was administered to the whole class rather than to specific pupils. Staff from intervention schools who were interviewed also highlighted that Year 2 pupils were engaging in longer and more regular handwriting practice sessions than usual. Interviewees reported that HHS typically replaced some of the usual literacy teaching for Year 2 intervention pupils during the eight-week intervention period. In most case-study schools, intervention delivery often took place during English or literacy lessons

when intervention pupils would usually receive handwriting practice sessions, spelling or phonics lessons, or participate in guided reading sessions, for example.

Usual practice for Year 2 classes in control schools

The baseline proforma showed that just under two thirds of Year 2 control schools intended to provide some form of additional literacy support to Year 2 pupils. Following a similar pattern to the planned practice of intervention schools (described above):

- the main form of additional literacy support planned for Year 2 control pupils was the use of specific **reading interventions or resources** (for example, Reading Recovery and Toe by Toe);
- the second most frequently cited form of planned support for Year 2 control pupils involved **specific literacy programmes** (for example, Read Write inc. and Lexia);
- the third most frequently cited planned additional literacy support for Year 2 control pupils was the use of **general literacy support** (for example, non-commercial or non-branded interventions) such as 'booster classes' and additional time with a TA to focus on individual literacy needs; and
- just two control schools said they would be delivering **specific handwriting interventions** (for example, Write from the Start) to Year 2 control pupils.

Just over a quarter of control schools responding to the baseline proforma reported that they would be using a mix of additional literacy support systems with their Year 2 pupils. In the majority of cases, this planned support was targeted at specific Year 2 pupils rather than the whole class and was usually provided by TAs.

Data from the business as usual log, completed during the same period as the intervention delivery, showed that in just over three quarters of schools Year 2 control pupils received literacy or handwriting support in addition to normal teaching. Almost two fifths of control schools that responded used a literacy intervention or programme with their Year 2 pupils and half used a specific handwriting intervention or programme. Almost all control schools allocated time specifically for handwriting practice in their Year 2 classes during this period. In most cases, the handwriting tasks and activities used within these sessions were not individualised. The majority of control schools delivered between two and three handwriting practice sessions to Year 2 pupils per week. Most of the sessions lasted between 10 and 30 minutes. These findings suggest that during the trial, Year 2 control pupils were spending similar amounts of time on handwriting as those involved in HHS.

Usual practice for Year 5 intervention and control group pupils (in intervention schools)

The baseline proforma showed that around three fifths of Year 5 intervention schools intended to provide some form of additional literacy support for their Year 5 pupils:[48]

- the main form of additional literacy support planned for Year 5 pupils involved **specific literacy programmes** (for example, Read Write inc. and Lexia);
- the second most frequently cited planned additional literacy support was the use of specific **reading interventions or resources** (for example, Catch Up® Literacy);
- the third most frequently cited form of planned support was **general literacy support** (for example, non-commercial or non-branded interventions) such as 'booster classes' or additional time with a TA to focus on individual literacy needs; and
- just four control group schools said they would be delivering **specific handwriting interventions** to target Year 5 control pupils (for example, Speed Up! and Paired Writing).

[48] Schools could name more than one programme or intervention each that they intended to use with their pupils.

Just under one fifth of intervention schools responding to the baseline proforma reported that they would be using a mix of additional literacy support systems with their Year 5 pupils. In the majority of cases, the planned additional literacy support for Year 5s would be targeted at specific pupils and provided by a combination of teachers and TAs.

- **Changes in practice for intervention Year 5 pupils.** Case-study interviewees highlighted that Year 5 intervention pupils were engaging in more frequent handwriting practice sessions than usual during the eight-week intervention period. The HHS sessions usually took place during non-'core' curriculum subject teaching time, assembly time, or time usually reserved for other support interventions. In a small number of case-study schools, Year 5 pupils were withdrawn from English lessons (including guided reading) to participate in HHS. These findings indicate that HHS was mostly provided in addition to usual literacy lessons for Year 5 intervention pupils.

- **Changes in practice for control Year 5 pupils**. Data from the business as usual log, completed during the same period as the intervention delivery, showed that one out of every ten Year 5 intervention schools used a literacy intervention or programme with their control pupils. One in every ten also used a specific handwriting intervention or programme with pupils in the control group during this time (although this contrasts with the findings from case study interviews with Year 5 staff who reported that very few control pupils received additional handwriting support in the trial period). Most schools provided time specifically for handwriting practice to Year 5 control pupils. This commonly involved two to three sessions per week lasting between 10 and 30 minutes. This finding suggests that during the trial, Year 5 control group pupils were spending similar amounts of time on handwriting practice to that involved in HHS.

Distinctive features of the intervention

The elements of HHS that differentiate it from usual practice were the same elements that were considered key to its success. The main distinctive feature of HHS identified by around a third of case-study interviewees is the formalised structure. For school staff, it provides a straightforward process to follow for teaching handwriting (a structure that some interviewees noted was absent from their teaching practice previously). For pupils, HHS offers consistent and regular handwriting sessions, and the repetitive structure means pupils quickly become familiar with the sequence of activities. Small numbers of interviewees flagged the self-evaluation element (a feature of the model-plan-evaluate process) as being a different element to their usual handwriting teaching, and perceived this to be particularly helpful in building pupils' confidence. Other effective elements of HHS compared to other usual literacy support or interventions identified by individual interviewees relate to the relative ease in which it could be administered by schools. This included the limited time required for preparation, the lack of additional materials or resources needed (other than photocopied activity sheets from the manual), its ability to fit with schools' existing writing policies, and its relevance and applicability to a broad range of pupils.

Perceived outcomes

This section explores the perceived outcomes of the intervention. It addresses perceived outcomes on pupils and school staff and draws entirely upon data from the case-study interviews.

Perceived impacts on pupils

Interviewees from all of the case-study schools reported positive benefits for pupils. School staff interviewed believed pupils in Year 2 and Year 5 had improved their handwriting skills and abilities over the eight-week intervention period, consistent with the theory of change. This was evidenced, for example, through staff observations of pupils' writing skills during the intervention sessions and in other lessons, as well as through improvements in the quality of written work in handwriting books over this time. Interviewees particularly highlighted improvements in pupils' fine motor skills (handgrip/control), writing accuracy and presentation, and writing fluency and speed (evidenced through their ability to write longer pieces, for example). Other perceived impacts on pupils involved a better understanding of the writing process and more positive attitudes towards handwriting, demonstrated through pupils' increased enthusiasm for, and enjoyment of, handwriting and increased pride in their written work. Year 2 and Year 5 pupils echoed some of these impacts themselves during case-study interviews with most reporting that HHS had helped their writing in some way. In some cases, however, staff interviewees felt the initial improvements they observed during the eight-week intervention period were not sustained in the longer term, which may have had an impact on the outcomes of the follow up testing.

We asked case-study school staff which pupils HHS works best for and around half of interviewees felt that HHS has a largely universal application and could be beneficial for most, if not all, pupils. A small number of Year 5 staff interviewed felt that either the wrong target pupils had been selected or that other Year 5 pupils may have benefited more. Some interviewees recognised that HHS was particularly effective for pupils who lack confidence in writing or find the process stressful, noting that the simplified approach to teaching writing builds pupils' self-esteem. Some interviewees supposed younger pupils would benefit more, before handwriting habits such as poor handgrip had become established, for example. A small number of interviewees felt that that the intervention worked less well with pupils with high-level behavioural needs and with particular characteristics and attitudes such as Year 5 pupils with a reluctance to receiving additional support, a dislike of being withdrawn from class, or who are less engaged with academic learning generally. Other individual interviewees reasoned that the intervention worked less well for pupils with an existing high standard of writing and those with very low writing abilities (who would typically require more specialist writing support).

Perceived impacts for schools

Interviews with case-study schools confirmed that delivering HHS has benefits for both teaching and support staff. Interviewees in over half of case-study schools described improved skills and abilities for teaching handwriting. Participating in the intervention had, for example, increased their subject knowledge and understanding of handwriting pedagogy, provided new ideas for handwriting activities, and offered an effective structure for their handwriting teaching. A small number of interviewees also reported an increased enjoyment of teaching handwriting and an awareness of, and improvement in, their own writing (while modelling to pupils, for example). Other impacts reported by individual case-study interviewees included increased confidence in teaching handwriting, increased awareness of writing across the curriculum, increased awareness of the need to prioritise the teaching of handwriting, improved understanding of individual pupil needs and improved relationships with pupils in Year 5 (where the intervention was typically delivered by TAs in small groups). These are distinct benefits to the intervention, though they are not necessarily specific to HHS or the topic of handwriting.

Cost

Training was the significant cost for schools involved in this intervention. In the Year 5 experiment, the intervention was more costly for schools on a per pupil basis because delivery took place in small groups. In the Year 5 intervention, one trained member of staff was delivering the intervention to significantly fewer pupils. Table 36 and Table 36 outline the financial costs of the experiements. In most schools training costs were lower for the Year 5 experiment as it tended to be TAs who attended the training and therefore supply costs for covering staff at the training have not been included for the second experiment. Sensitivity analysis was completed to assess the difference in per pupil costs between whether a TA or a teacher attended the training. In the Year 2 experiment, a TA attending the training instead of a teacher reduced the per pupil cost (over three years) by £1, as shown in last column of Table 36. In the Year 5 experiment, a teacher attending the training instead of a TA increased the per pupil cost by £8 to £180 (see Table 36).

Table 36: Cost of delivery—Year 2 experiment

		Start-up or Recurring?	Quantity required	Price per Unit Required	Cost per year	Total (with teacher attending training)	Total (with TA attending training)
Personnel for training*	Supply cover for TA attending training	Start-up	1 staff member (mode) for 0.5 day	£0	£0		£0
	Supply cover for teacher attending training	Start-up	1 staff member (mode)	£200 per day	£100	£100	

			for 0.5 day				
Training and programme costs	Staff training session, training resources, intervention handbook, pupil resource booklets	Start-up	1	£2000 per school	£2,000	£2,000	£2,000
	Subsistence/travel costs associated with training attendance	Start-up	1 (mode)	£82	£82	£82	£82
Facilities, equipment and materials **	Resources - photocopying, paper, pencils (Year 2)	Recurring	1	£7.50	£7.50	£22.50	£22.50
Total cost per school for 3 years						£2,205	£2,105
Total cost per pupil per year***						£27	£26

* The calculations in this table are based on 1 teacher per school attending the training. Data from the training registers shows that 79% of Y2 training attendees were teachers. The estimated cost to schools for supply teacher cover is £200 per day. Training attendance data shows schools modal number of staff attending training per school is 1 per year group (26/46 schools sent 1 teacher, 16/46 sent 2). Calculations are therefore based on 1 Y2 class undertaking the intervention per school.
** Based on midpoint of modal response from Year 2 logs referring to 'photocopying/printing' or 'work files' (n = 17) to question about costs.
*** Based on average primary class size of 27.1 (DfE, 2019) and assuming one Year 2 class per school undertaking the intervention.

Table 35: Preparation and delivery time—Year 2 experiment

		Year 1		Year 2		Year 3	
		Number of teachers	Number of hours per week	Number of teachers	Number of hours per week	Number of teachers	Number of hours per week
Preparation*	Teacher	1 (mode)	0.25	1 (mode)	0.25	1 (mode)	0.25
Delivery	Teacher	1 (mode)	1.5	1 (mode)	1.5 hours	1 (mode)	1.5 hours

*Based on midpoint of the modal group of modal response from Year 2 logs in response to question about extra time spent on intervention (61% responded 'less than 30 minutes per week', n = 31).

Table 36: Cost of delivery—Year 5 experiment

		Start-up or Recurring?	Quantity required	Price per Unit Required	Cost per year	Total (with teacher attending training)	Total (with TA attending training)
Personnel for training *	Supply cover for TA attending training	Start-up	1 staff member	£0	£0		£0

			(mode) for 0.5 day				
	Supply cover for teacher attending training	Start-up	1 staff member (mode) for 0.5 day	£200 per day	£100	£100	
Training and programme costs	Staff training session, training resources, intervention handbook, pupil resource booklets	Start-up	1	£2000 per school	£2,000	£2,000	£2,000
	Subsistence/travel costs associated with training attendance	Start-up	1 (mode)	£82	£82	£82	£82
Facilities, equipment and materials **	Resources - photocopying, paper, pencils (Year 5)	Recurring	1	£2.50	£2.50	£7.50	£7.50
Total cost per school for 3 years						£2,190	£2,090
Total cost per pupil per year***						£180	£172

* The calculations in this table are based on one TA per school attending the training. Data from the training registers shows that 85% of Year 5 training attendees were TAs. We have assumed schools do not use supply cover for TAs out of school at training. Training attendance data shows schools modal number of staff attending training per school is one per year group (35/41 schools sent one, 5/41 sent two). Calculations are therefore based on one Year 5 intervention group per school.
** Based on the midpoint of modal response from Year 5 logs referring to 'photocopying/printing' (n = 10) to question about costs.
*** Based on average number in a teaching group for the Year 5 experiment using numbers at randomisation (186/46 = 4.04).

Table 37: Preparation and delivery time—Year 5 experiment

		Year 1		Year 2		Year 3	
		Number of teachers	Number of hours per week	Number of teachers	Number of hours per week	Number of teachers	Number of hours per week
Preparation *	Teaching assistant	1 (mode)	0.25	1 (mode)	0.25	1 (mode)	0.25
Delivery	Teaching assistant	1 (mode)	1.5	1 (mode)	1.5 hours	1 (mode)	1.5 hours

*Based on midpoint of the modal group of modal response from Year 5 logs in response to question about extra time spent on intervention (61% responded 'less than 30 minutes per week', n = 34).

Conclusion

Key Conclusions	
1.	Children in the Helping Handwriting Shine schools who were in Year 2 and experienced the universal intervention made no additional progress, on average, in their overall writing ability compared to children in the control group. The range of possible impacts for the universal programme in Year 2 includes small negative effects of two months' less progress and small positive effects of up to one month's progress. This result has a high security rating.
2.	Children in the Helping Handwriting Shine schools who were in Year 5 and experienced the targeted intervention made the equivalent of two additional months' progress in their overall writing ability compared to children in the control group. The range of possible impacts for the targeted programme in Year 5 include small negative effects of one month less progress and moderate positive effects of up to four months' progress. This result has a moderate to high security rating.
3.	Children in the Helping Handwriting Shine schools, either in Year 2 or Year 5, made, on average, no additional progress in writing composition. This result may have lower security than the overall findings. Exploratory analysis also suggests that it is unlikely the Helping Handwriting Shine programme increased or decreased children's handwriting speed.
4.	Year 2 children in Helping Handwriting Shine schools who were eligible for Free School Meals (FSM) made, on average, no additional progress in their overall writing ability compared to similar children in the control group. Year 5 children in Helping Handwriting Shine schools who were eligible for FSM made, on average, two additional months progress in their overall writing ability compared to similar children in the control group. These analyses are exploratory and have a lower security than the headline findings given the smaller subsample of children included (Y2 FSM n = 774 and Y5 FSM n = 91)
5.	Adherence to the eight-week programme was, on average, medium to high. There were some potential limitations with the delivery logs meaning these findings should be viewed with caution. Staff and pupils viewed the programme positively, noticing improvements in children's handwriting during the eight-week intervention. Staff were of the opinion that the programme had improved their ability to teach handwriting and that it was relatively easy to implement.

Impact evaluation and IPE integration

This study set out to investigate whether a handwriting intervention could be effectively delivered by teachers and teaching assistants in schools, as opposed to being delivered by occupational therapists in clinics. It also set out to investigate the impact of the intervention on writing composition. The intervention had no effect on either the target cohort's handwriting speed nor their writing composition, observed at the end of the trial, for either the Year 2 or Year 5 experiment. There was also no effect observed in ideal conditions for either cohort. In this section, we consider these findings in light of the theory of change (ToC), as well as the strengths and limitations of the trial.

Evidence supporting the theory of change (ToC)

The ToC for this intervention is based around the key assumption that 'once a child is automatic at handwriting they can become more fluent at technical writing', with the hypothesis that this will lead to 'freed up cognitive capacity enabling better writing composition'. Dosage and delivery requirements are stipulated that, when adhered to, will lead to this outcome. It is perhaps worth breaking this down more clearly to reflect the two components to this hypothesis, and how they are linked.

> Component 1: upon receiving the intervention, a child becomes better at handwriting—they are better able to produce fast, accurate and legible handwriting.
>
> Component 2: *as a result of 1*, the child's writing composition improves.

The outcomes of this trial show that, despite generally good fidelity to the dosage and delivery requirements stipulated, there is no evidence of a child's handwriting, as measured by the Handwriting Speed Test after the gap period, improving on receipt of the intervention. There are caveats to this, discussed later, but it must be reflected that there is no evidence of the success of component 1 for either Year 2 or Year 5 (as measured using secondary outcome 1, the Handwriting Speed Test).

Because of this, it is difficult, if not impossible, to draw any conclusions about the potential link between component 1 and component 2. With no measurable difference in speed seen in secondary outcome 1, any change seen in writing

composition cannot be attributed to this, and the link between components 1 and 2 in the ToC is broken. There was also no significant change seen in writing composition.

This break in the link of the ToC prompts consideration of the concept of component 1: whether the evidence at commencement of the trial was strong enough for this to be a fundamental element of the ToC, underpinning the success of the trial. As identified in the background section, there is evidence of success in clinical intervention via an occupational therapist; this is on a one-to-one basis, however. Other handwriting programmes, such as Write from the Start and Letter Join, or interventions such as Rapid Writing, are reasonably well established within UK schools but have not been subject to robust evaluation. The HHS intervention, therefore, is a relative front-runner; it also uses well-evidenced techniques and concepts such as task-oriented approaches (Smits-Engelsman et al., 2013; Preston et al., 2017) and the identify-try-reflect-refine approach (essentially a metacognitive structure: Smits-Engelsman et al., 2013; Pless and Carlson, 2000). We may conclude that the evidence of the HHS intervention potentially improving handwriting was strong, but as it is unprecedented in its style of delivery and intention, it may have been more appropriate to test the two components separately in one or more RCTs, as opposed to making component 1 a foundation to component 2 in one experiment.

It is also worth considering that the results of component 1 may have been different had measurement (secondary outcome measure 2) been conducted at the end of the eight-week intervention period, as opposed to following an extended 'gap' period; this was reflected on during some case-study interviews, where staff felt the initial improvements pupils had made during the formal intervention period were not sustained over the longer term. Teachers, in fact, perceived a high impact from the intervention on manual control, fluency and speed as well as enthusiasm and pride in their own work, which leads us to consider the embedding period and its potential impact on outcomes. Furthermore, our business-as-usual logs revealed that controls were spending similar amounts of time on handwriting interventions as HHS pupils in both experiments so our results suggest that these activities might be 'as good as' HHS.

As the primary outcome results for all pupils in the trial were mirrored for FSM-eligible pupils, there is no reason to think that this break in the ToC was not also present for these pupils.

Embedding period

The requirement for a gap period was suggested by the evaluator at outset of the research. The purpose of the gap period was to give schools the opportunity to embed the intervention into their normal practice and allow change to take effect. It was discussed at outset that an ideal design would allow measurement both immediately after the intervention and then again after a period of time had elapsed. However, the burden this would place on schools would be significant, and the cost prohibitive, so endpoint testing was scheduled for later in the year. There are two complications with the embedding period that were not foreseen.

Understanding the nature, extent, and effect of embedding practice

In the documents shared with schools for recruitment, it was shown in the school timelines that intervention schools would embed the approaches, as per this extract from the School Information Sheet:

Spring term 2019: Intervention schools embed approaches

- Staff complete online logs
- Complete follow up tasks to embed approaches

The same wording was included in the timeline within the Memorandum of Understanding; however, under the 'specific expectations' of intervention schools within the Memorandum of Understanding, the embedding period is not listed. Listed under 'specific requirements' are attending the training, delivering the intervention over the eight weeks, completing intervention delivery logs, and conducting the baseline and endpoint testing; they did not explicitly sign up to embed the intervention into their normal practice. This may have resulted in a lack of engagement during this period.

Conversely, though, the developer team introduced the concept of the gap period—its purpose and how it would be supported—in the initial training at the commencement of the intervention. A half-day training session was provided to schools to help transition into normal teaching and to support schools in creating an action plan for this transition and for the following period until the end of the school year. Communication between the developer team and schools was weekly during this period. Additionally, schools were invited to design a gap or extension task as part of a competition;

these were shared in newsletters with schools. Eleven newsletters were sent between 1 March and 24 May 2019; the first ten newsletters contained four gap tasks ansd the last newsletter contained eight (48 in total). In a final additional competition, all gap and extension tasks were compiled into a compendium and schools were invited to vote on the best submissions, with the two schools with the highest votes winning a prize delivered by the developer team. Clearly, during this period, there was extensive engagement from schools, but monitoring this engagement, and what activity took place during this period, was challenging. Schools were required to state whether they were embedding the intervention, and how regularly, using a gap period monitoring log, which was due to be completed once each month during the gap period and submitted at the end of the period. However, the response rate for these logs was very low. Of the 18 schools that did return a log, only six reported that they were embedding the programme into their usual practice in any significant way. Interviews conducted towards the end of the trial indicated some embedding took place during this period, including using some of the tasks, doing shorter sessions less frequently, or taking specific elements of the intervention that the teacher felt were working well or the pupils enjoyed. However, the metacognitive element was rarely retained.

Reflecting on the communications with schools during this period, there was, consciously, a greater focus on requiring schools to (a) make arrangements for post testing and (b) return the intervention delivery logs than there was on requiring the completion and return of the gap period monitoring log as these were felt by the evaluators to be higher priority for the evaluation. This was also, in part, to ensure the evaluation team were not overburdening schools with too many 'chasing' emails during this period. Balancing these requirements is difficult for any evaluator given that element (a) is fundamental to all outcomes and (b) is fundamental to the CACE analysis. However, on reflection of the potential ambiguity of the embedding period for schools and how fundamental the embedding period is in understanding the connection between components 1 and 2, it may have been prudent to give an equal focus on these *as well* as collection of the gap period monitoring log, with an acknowledgment across the teams that this could lead (at worst) to further attrition.

Year 5

The Year 5 experiment was randomised at pupil level within schools that were allocated to the intervention arm of the Year 2 experiment. In practice, this meant it was crucial to avoid contamination between the Year 5 intervention and control pupils, which was carefully managed by the evaluator and developer teams. It was suggested that, unless impossible to accommodate, intervention pupils should be removed from normal lessons and taught separately by a member of staff who did not, or would not, teach the control pupils. Contamination was highlighted within the training sessions, printouts with rules were provided to schools, and teachers were found to be well aware of the requirements around avoiding contamination within the IPE.

However, this had a predictable effect on the Year 5 experiment: once the formal intervention period was over, intervention pupils went back to normal classes and were taught normal lessons by untrained staff without any embedding of intervention techniques (so as not to contaminate the control pupils). Essentially, embedding was not possible within the Year 5 experiment and so Year 5 pupils received only the eight weeks of the formal intervention without any further embedding. Interviews demonstrated that in some schools, the staff members who had delivered the Year 5 intervention were immediately re-tasked on other projects or support for other pupils and were not able to engage with the HHS intervention pupils again at all after the formal intervention period.

This would go some way to suggest that future research for the Year 5 age group may be more appropriately designed to integrate the ability to embed, perhaps as a feature of a cluster-grouped trial as opposed to this trial's in-school design.

Comparative judgment

To the authors' knowledge, this is the first UK trial to use comparative judgment (CJ) as the assessment method for a randomised controlled trial in education and as such it is worth reflecting upon any impact this may have had on delivery, outcomes, and relative success of the trial. The evaluator team were able to use the outcome data produced by the comparative judgment process in much the same way as 'normal' assessment data provided by criteria-linked assessment methods. By the end of the six-week period with all judging complete, reliability of the CJ outcomes sat at over 0.92 (scale separation index).

The benefits of CJ were significant, primarily because it allowed us to mark a much larger sample than would have been possible with traditional criterion-based marking. On a sample of this size, traditional marking would have been prohibitively costly, and taken much longer, with many more markers to manage and standardise, increasing the burden of marker management to ensure high reliability. NFER was able to pilot the use of the NMM platform before it was used

for the baseline tasks, which proved important in realising and solving logistical interruptions such as unreadable QR codes.

Schools were provided with feedback, as is normal practice for NFER trials, using their endpoint scores. As the scores are relative to the context of the schools that participated in the trial and not anchored to other standards, NFER divided the participating schools into five attainment quintiles according to their average comparative judgment scores (top quintile, mid-high quintile, middle quintile, mid-low quintile, and low quintile). Following a similar approach, NFER also grouped Year 2 pupils who participated in the trial into quintile groups according to the comparative judgement scores they got on the test. Each score quintile comprises approximately 20% of the 3,472 Year 2 pupils who were assessed in the trial. Lastly, the rank was provided for each pupil who took the test within the school. As such NFER were able to provide useful, comprehensible feedback at a school and pupil level.

Limitations and lessons learned

The limitations within this research mostly concern the design aspects around the key issues described above: the within-school design for Year 5 may not have accurately represented the intervention as, to minimise contamination, teachers were not expected to complete gap tasks with participating pupils and were less able to embed practice. More generally, ambiguity around the definition and understanding of the embedding period, also discussed above, may have led to unpredictable variation within post testing since it occurred after the embedding period.

Some pupil-level attrition occurred across the two experiments (Year 2, 11%; Year 5, 18%), balanced across the trial arms. All judges and markers were blinded to condition, and while teachers were not blinded when administering the test at baseline, independent test administrators delivered the test at endpoint so bias here was mitigated.

The number of case studies and interviews in the process evaluation means that outcomes from this element of the evaluation are not generalizable. All schools involved in the process evaluation volunteered for the activity, which introduces a positive bias; however, their insight did provide in-depth evidence for individual practice as opposed to breadth and potential explanations for trends seen in the impact evaluation.

All logs disseminated to schools would have benefitted from a pilot phase; this may have picked up on the burden of completion. A potential limitation of the delivery log for Year 5 pupils was identified at the analysis stage. In most cases, there appeared to be no variation in the information given for pupils within each Year 5 group. This could reflect that very little tailoring or adaptation of the intervention was required for individual pupils in these groups. Alternatively, it could indicate that in order to save time on completion, staff copied the information for each pupil, rather than recording on an individual basis. Indeed, while school staff were encouraged to complete logs after each session, the evaluators were made aware that some schools completed them retrospectively (for example, on a weekly, or longer, basis). This may have affected their ability to recall small variations in delivery on a pupil-by-pupil basis. The findings of the Year 5 delivery log should therefore be viewed with caution. The same limitations with the completion of the Year 5 post programme logs (as the delivery logs) were identified at the analysis stage. There was little variation in the information given for each pupil within each Year 5 group. Therefore, these findings should be viewed with caution.

The Handwriting Speed Test would also have benefitted from piloting. This test ended up representing the key measure for this evaluation since it assessed the first stage of the ToC as being ineffective. As the analysis model had to be changed to reflect the data from the test, we cannot be sure whether it was of adequate statistical power. Some further work on simulation, which was out-of-scope for this evaluation, may well resolve this. However, the histograms in Appendices D1 and D2 provide adequate evidence that children's handwriting speed was not changed in any appreciable way as a result of this intervention (though some pupils may not have written as much as they could have if they stopped at complete sentences).

Future research and publications

It may be desirable for future research to focus on the two components identified here, separately, to provide robust evidence of a link between the intervention and improved handwriting speed and fluency, before further evidence is sought of a link between this and a further improvement in writing composition.

References

Andrade, J., Styles, B. and Stone, G. (2019) Helping Handwriting Shine Statistical Analysis Plan Available at: **HelpHandwritingShine_SAP_vFinal_10.05.2019.pdf (d2tic4wvo1iusb.cloudfront.net)** [5th Jan 2022]

Berninger, V. W. and Graham, S. (1998) 'Language by Hand: A Synthesis of a Decade of Research on Handwriting', *Handwriting Review*, 12, pp. 11–25. DOI 10.1177/002221940203500104.

American Psychiatric Association (2013) *American Psychiatric Association Diagnostic and Statistical Manual of Mental Disorders* (5th edn), Arlington, Virginia.

Christensen, R. H. B (2019) 'Ordinal-Regression Models for Ordinal Data', *R package version 2019*, 12-10.

Department for Education (2019) 'Schools, Pupils and Their Characteristics: January 2019', DfE: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/812539/Schools_Pupils_and_their_Characteristics_2019_Main_Text.pdf

Dunford, C. and Richards, S. (2003) 'Doubly Disadvantaged: Report of a Survey on Waiting Lists and Waiting Times for Occupational Therapy Services for Children with Developmental Coordination Disorder' Project Report, College of Occupational Therapists: https://ray.yorksj.ac.uk/id/eprint/50/1/doubly-disadvantaged.pdf

Dunsmuir, S., Kyriacou, M., Batuwitage, S., Hinson, E., Ingram, V. and O'Sullivan, S. (2015) 'An Evaluation of the Writing Assessment Measure for Children's Narrative Writing', *Assessing Writing*, 23, pp. 1-18. DOI: 10.1016/j.asw.2014.08.001.

EEF (2013) 'Pre-Testing in EEF Evaluations', Education Endowment Foundation: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol_or_SAP/Pre-testing_paper.pdf

EEF (2015) 'Intra-cluster Correlation Coefficients', Education Endowment Foundation: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol_or_SAP/ICC_2015.pdf

EEF (2018) 'Statistical analysis guidance for EEF evaluations, March 2018', Education Endowment Foundation. https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol_or_SAP/EEF_statistical_analysis_guidance_2018.pdf

Evidence4impact (2018) 'Effective Educational Intervention Database': https://www.evidence4impact.org.uk/

Hoy, M. M. P., Egan, M. Y. and Feder, K. P. (2011) 'A Systematic Review of Interventions to Improve Handwriting', *Canadian Journal of occupational therapy,* 78, 1, 3. DOI: 10.2182/cjot.2011

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R. and Kerr, K. (2016) 'Implementation and Process Evaluations (IPE) for Interventions in Education Settings: An Introductory Handbook', Education Endowment Foundation:
https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/IPE_Handbook.pdf

Hunter, D. R. (2004) 'MM Algorithms for Generalized Bradley-Terry Models', *Annals of Statistics*, 32, 1, pp. 384–406: http://personal.psu.edu/drh20/papers/bt.pdf

Angrist, J. D. and Imbens, G. W. (1995) 'Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity', *Journal of the American Statistical Association*, 90, 430, pp. 431–442. DOI: 10.1080/01621459.1995.10476535

Kent, S. C. and Wanzek, J. (2016) 'The Relationship between Component Skills and Writing Quality and Production Across Developmental Levels: A Meta-Analysis of the Last 25 Years', *Review of Educational Research*, 86, 2, pp. 570-601. DOI: 10.3102/0034654315619491

Kish, L. (1965) *Survey Sampling*, New York: J. Wiley and Sons.

Liu, T., Hoffmann, C. and Hamilton, M. (2015) 'Motor Skill Performance by Low SES Preschool and Typically Developing Children on the PDMS-2', *Early Childhood Education*, 45, 1, pp. 53–60. DOI: 10.1007/s1064301507559

Lortie-Forgues, H. (2017) 'What Can We Learn from RCTs in Education?', paper presented at the 'RCTs in the Social Sciences' Conference, University of York, September 2017.

McCarney, D., Peters, L., Jackson, S., Thomas, M., Kirby, A. (2013) 'Does Poor Handwriting Conceal Literacy Potential in Primary School Children?' *International Journal of Disability, Development, and Education,* 60, 2, pp. 105–118.

McCutchen, D. (1996) 'A Capacity Theory of Writing: Working Memory in Composition', *Educational Psychology Review*, 8, 3, pp. 299–325.

McKnight, P. E., McKnight, K. M., Sidani, S. and Figueredo, A. J. (2007) *Missing Data: A Gentle Introduction*, New York: Guildford Press.

Medwell, J., Strand, S. and Wray, D. (2009) 'The Links Between Handwriting and Composing for Y6 Children', *Cambridge Journal of Education,* 39, 3, pp. 329–344. DOI: 10.1080/03057640903103728

Pless, M. and Carlsson, M. (2000) 'Effects of Motor Skill Intervention on Developmental Coordination Disorder: A Meta-Analysis', *Adaptive Physical Activity Quarterley,* 17, pp. 381–401.

Pollitt, A. (2012) 'The Method of Adaptive Comparative Judgement', *Assessment in Education: Principles Policy and Practice*, 19, 3 (whole issue). DOI: 10.1080/0969594X.2012.665354

Preston, N. et al. (2017) 'A Systematic Review of High Quality Randomized Controlled Trials Investigating Motor Skills Programmes for Children with Developmental Coordination Disorder', *Clinical Rehabilitation*, 31, 7, pp. 857–870.

Sanders, M. and Ni Chonaire, A. (2015) '"Powered to Detect Small Effect Sizes": You Keep Saying That. I Do Not Think It Means What You Think It Means' (Working Paper No. 15/337): http://www.bris.ac.uk/media-library/sites/cmpo/documents/WP15337_Web_Version.pdf

Santangelo, T. and Graham, S. (2016) 'A Comprehensive Meta-analysis of Handwriting Instruction', *Educational Psychology Review,* 28, pp. 225–265. DOI: 10.1007/s10648-015-9335-1

Shire, K.A., Atkinson, J., Williams, E.A., Pickavance, J., Magallón, S., Hill, J.B.H., Waterman, A.H., Sugden, D.A., & Mon-Williams. M. (in press). Developing and implementing a school-led motor intervention for children with handwriting difficulties. *Journal of Occupational Therapy, Schools, & Early Intervention*

Smits-Engelsman, B. C. M. et al. (2013) 'Efficacy of Interventions to Improve Motor Performance in Children with Developmental Coordination Disorder: A Combined Systematic Review and Meta-Analysis', *Developmental Medicine and Child Neurology,* 55, 3, pp. 229–37. DOI: 10.1111/dmcn.12008

Snijders, T. A.B. (2005) 'Power and Sample Size in Multilevel Linear Models', in B. S. Everitt and D. C. Howell (eds), *Encyclopedia of Statistics in Behavioral Science*, Chicester: Wiley.

Torgerson, D., Torgerson, C., Mitchell, N., Buckley, H., Ainsworth, H., Heaps, C. and Jefferson, L. (2014a) 'Grammar for Writing: Evaluation Report and Executive Summary', Education Endowment Foundation: https://www.researchgate.net/publication/266967404_Grammar_for_Writing_Evaluation_Report_and_Executive_Summary

Torgerson, D., Torgerson, C., Mitchell, N., Buckley, H., Ainsworth, H., Heaps, C. and Jefferson, L. (2014b) 'Improving Writing Quality: Evaluation Report and Executive Summary', Education Endowment Foundation: https://files.eric.ed.gov/fulltext/ED581140.pdf

Van Buuren, S. and Groothuis-Oudshoorn, K. (2011) 'mice: Multivariate Imputation by Chained Equations in R', Journal of Statistical Software, 45, 3, pp. 1–67.

Wallen, M., Bonney, M. A. and Lennox, L. (2006) 'The Handwriting Speed Test', *Australian Occupational Therapy Journal*, 60, 5, pp. 141–144.

Wallen, M., Duff, S., Goyen, T-A. and Froude, E. (2013) 'Respecting the Evidence: Responsible Assessment and Effective Intervention for Children with Handwriting Difficulties', *Australian Occupational Therapy Journal*, 60, 5, pp. 366–369. DOI: 10.1111/1440-1630.12045

Appendix A: EEF cost rating

Appendix Table A1: Cost Rating

Cost rating	Description
£ £ £ £ £	*Very low:* less than £80 per pupil per year.
£ £ £ £ £	*Low:* up to about £200 per pupil per year.
£ £ £ £ £	*Moderate:* up to about £700 per pupil per year.
£ £ £ £ £	*High:* up to £1,200 per pupil per year.
£ £ £ £ £	*Very high:* over £1,200 per pupil per year.

Appendix B: Security classification of trial findings

OUTCOME: *Writing Assessment Measure – Comparative Judgement (WAM_CJ), Y2 trial*

Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	**MDES**	**Attrition**					
5	Randomised design	<= 0.2	0-10%					
4	Design for comparison that considers some type of selection on unobservable characteristics (e.g. RDD, Diff-in-Diffs, Matched Diff-in-Diffs)	0.21 - 0.29	11-20%	4				4
3	Design for comparison that considers selection on all relevant observable confounders (e.g. Matching or Regression Analysis with variables descriptive of the selection mechanism)	0.30 - 0.39	21-30%			Adjustment for threats to internal validity **[0]**		
2	Design for comparison that considers selection only on some relevant confounders	0.40 - 0.49	31-40%					
1	Design for comparison that does not consider selection on any relevant confounders	0.50 - 0.59	41-50%					
0	No comparator	>=0.6	>50%					

Threats to validity	Threat to internal validity?	Comments
Threat 1: Confounding	Moderate	.10 to .13 imbalance in writing assessment measure at baseline in favour of the control group, but well controlled in analysis.
Threat 2: Concurrent Interventions	Low	Concurrent interventions are measured well, clearly reported, and adopted to a similar extent by both groups.
Threat 3: Experimental effects	Low	Minimal evidence of experimental effects or contamination.
Threat 4: Implementation fidelity	Moderate	Reported fidelity to the intervention is moderate to high, but issues with delivery and post-programme log data reduce the ability to interpret fidelity accurately.[49]
Threat 5: Missing Data	Low	Total missing data is moderate (11%) but there is no evidence of differential attrition and results after imputation are very similar to complete case analysis.
Threat 6: Measurement of Outcomes	Low	Outcome testing was conducted and marked independently and blinded to treatment allocation. The instrument and assessment methodology were selected to give an overall measurement of writing quality with high reliability.
Threat 7: Selective reporting	Low	No evidence of selective reporting when comparing the report with protocol and SAP. Deviations are clearly noted and justified.

[49] The add-on period where implementation is embedded is also not clearly defined, making it difficult to assess fidelity of this aspect, but this is not a core component of the intervention so is not considered here.

- **Initial padlock score:** [4] Padlocks – Well-designed RCT with low MDES at randomisation (0.18) and moderate attrition at point of analysis (11%)
- **Reason for adjustment for threats to validity**: [0] Padlocks – No evidence of severe threats to internal validity
- **Final padlock score:** Initial score adjusted for threats to validity = [4] Padlocks

OUTCOME: *WAM_CJ, Y5 trial*

Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	**MDES**	**Attrition**					
5	Randomised design	<= 0.2	0-10%					
4	Design for comparison that considers some type of selection on unobservable characteristics (e.g. RDD, Diff-in-Diffs, Matched Diff-in-Diffs)	0.21 - 0.29	11-20%	4				
3	Design for comparison that considers selection on all relevant observable confounders (e.g. Matching or Regression Analysis with variables descriptive of the selection mechanism)	0.30 - 0.39	21-30%			Adjustment for threats to internal validity **[-1]**		3
2	Design for comparison that considers selection only on some relevant confounders	0.40 - 0.49	31-40%					
1	Design for comparison that does not consider selection on any relevant confounders	0.50 - 0.59	41-50%					
0	No comparator	>=0.6	>50%					

Threats to validity	Threat to internal validity?	Comments
Threat 1: Confounding	Moderate	.08 to .11 imbalance in writing assessment measure at baseline in favour of the intervention group, but well controlled in analysis.
Threat 2: Concurrent Interventions	Low	Concurrent interventions are measured well, clearly reported, and adopted to a similar extent by both groups.
Threat 3: Experimental effects	Moderate/High	Although measures were taken to minimise this, the within-school design led to contamination reported in interviews with both teachers and developers. The full extent of this is unknown and is a significant risk.
Threat 4: Implementation fidelity	Moderate/High	Reported fidelity to the intervention is moderate to high, but issues with delivery and post-programme log data reduce the ability to interpret fidelity accurately. Additionally, to minimise contamination, some aspects of the intervention – e.g. gap tasks between HHS sessions – were not delivered as intended.[50]
Threat 5: Missing Data	Low	Total missing data is moderate (16%) but there is no evidence of differential attrition and results after imputation are very similar to complete case analysis.
Threat 6: Measurement of Outcomes	Low	Outcome testing was conducted and marked independently and blinded to treatment allocation. The instrument and assessment methodology were selected to give an overall measurement of writing quality with high reliability.
Threat 7: Selective reporting	Low	No evidence of selective reporting when comparing the report with protocol and SAP. Deviations are clearly noted and justified.

[50] Teachers in the Y5 experiment were also less able to embed practice in the post-intervention period so as not to contaminate control pupils. As the embedding period is not a core component of the programme, we do not consider it in the assessment of fidelity, but this may have led to unpredictable variation within post-testing.

- **Initial padlock score:** [4] Padlocks – Multisite RCT with moderate MDES (0.23) and moderate attrition at point of analysis (18%)
- **Reason for adjustment for threats to validity**: [-1] Padlocks – Subtract 1 padlock for the validity threat posed by contamination
- **Final padlock score:** Initial score adjusted for threats to validity = [3] Padlocks

Appendix C1: Results of the first stage of the analysis in the presence of non-compliance

Appendix table C1.1: Compliance 1 (school level) regressed on the instrumental variable (treatment allocation) and covariates included on the second stage (baseline WAM_CJ and region) for Year 2 Experiment*

	Raw coefficient	Standard Error	p-value
(Intercept)	0.06	0.06	0.32
Treatment**	11.58 (intervention)	0.05 (intervention)	< 0.00 (intervention)
Baseline WAM_CJ	-0.02	0.01	0.01
Region***	-0.10 (Leeds) -0.029 (Newcastle) 0.24 (Sheffield)	0.07 (Leeds) 0.07 (Newcastle) 0.08 (Sheffield)	0.15 (Leeds) <0.00 (Newcastle) <0.00 (Sheffield)

. * Adjusted R^2=0.953
** baseline category is control, *** baseline category is Darlington

Appendix table C1.2: Compliance 2 (school level) regressed on the instrumental variable (treatment allocation) and covariates included on the second stage (baseline WAM_CJ and region) for Year 2 Experiment*

	Raw coefficient	Standard Error	p-value
(Intercept)	-0.02	0.01	0.07
Treatment**	0.90 (intervention)	0.01 (intervention)	<0.00 (intervention)
Baseline WAM_CJ	0.00	0.00	0.08
Region***	-0.01 (Leeds) 0.01 (Newcastle) 0.06 (Sheffield)	0.01 (Leeds) 0.01 (Newcastle) 0.01 (Sheffield)	0.64 (Leeds) 0.22 (Newcastle) <0.00 (Sheffield)

. * Adjusted R^2=0.839
** baseline category is control, *** baseline category is Darlington

Appendix table C1.3: Compliance 1 (pupil level) regressed on the instrumental variable (treatment allocation) and covariates included on the second stage (baseline WAM_CJ FSM (reported), and school*) for Year 5 Experiment**.

	Raw coefficient	Standard Error	p-value
(Intercept)	0.27	0.44	0.54
Treatment**	10.70 (intervention)	0.18 (intervention)	< 0.00 (intervention)
Baseline WAM_CJ	0.02	0.05	0.73
FSM (reported)[iv]	-0.11 (FSM)	0.23 (FSM)	0.63 (FSM)

. * The information on school was not displayed on the table due to the large number of categories/levels of the variable (44).
** Adjusted R^2=0.925
*** baseline category is control, [iv] baseline category is non-FSM

Appendix table C1.4: Compliance 2 (pupil level) regressed on the instrumental variable (treatment allocation) and covariates included on the second stage (baseline WAM_CJ FSM (reported), and school*) for Year 5 Experiment**.

	Raw coefficient	Standard Error	p-value
(Intercept)	0.00	0.06	1.00
Treatment**	0.83 (intervention)	0.03 (intervention)	< 0.00 (intervention)
Baseline WAM_CJ	0.00	0.01	0.62
FSM (reported)[iv]	-0.01 (FSM)	0.03 (FSM)	0.85 (FSM)

* The information on school was not displayed on the table due to the large number of categories/levels of the variable (44).
** Adjusted R^2=0.814
*** baseline category is control, [iv] baseline category is non-FSM

Appendix table C1.5: Correlations between treatment allocation and the endogenous compliance variables and their associated F-test for Year 2 Experiment

	Correlation with instrumental variable (treatment)	F statistic for the weak-instruments test	p-value
Compliance 1 (school level)	0.98	61,732.66	< 0.00
Compliance 2 (school level)	0.91	15,958.37	< 0.00

Appendix table C1.6: Correlations between treatment allocation and the endogenous compliance variables and their associated F- test for Year 5 Experiment)

	Correlation with instrumental variable (treatment)	F statistic for the weak-instruments test	p-value
Compliance 1 (pupil level)	0.94	3,379.94	< 0.00
Compliance 2 (pupil level)	0.83	1,085.72	< 0.00

Appendix C2: Effect size and ICC estimation

Appendix table C2.1: Effect size estimation for Year 2 Experiment

			Intervention group		Control group			
Outcome	Unadjusted differences in means	Adjusted differences in means	n (missing)	Variance of outcome	n (missing)	Variance of outcome	Total variance from a model without covariates	Population variance (if applicable)
WAM_CJ Year 2	-0.275	-0.043	1679 (196)	6.724	1765 (214)	6.599	6.660	
WAM_CR Year 2	-0.43	-0.169	230 (30)	14.51	213 (42)	12.835	13.705	
WAM_CJ (CACE Comp 1) Year 2	-0.275	-0.008	1679 (196)	6.724	1765 (214)	6.599	6.660	
WAM_CJ (CACE Comp 2) Year 2	-0.275	-0.1	1679 (196)	6.724	1765 (214)	6.599	6.660	

Appendix table C2.2: Effect size estimation Year 5 Experiment

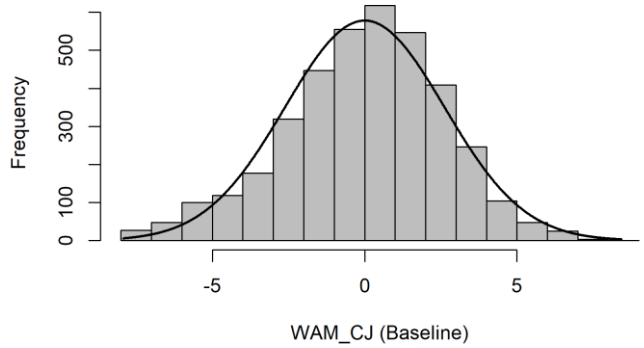
			Intervention group		Control group			
Outcome	Unadjusted differences in means	Adjusted differences in means	n (missing)	Variance of outcome	n (missing)	Variance of outcome	Pooled variance	Population variance (if applicable)
WAM_CJ Year 5	0.375	0.224	167 (27)	3.692	162 (24)	3.357	3.527	
WAM_CR Year 5	0.1	-0.152	167 (27)	11.578	162 (24)	12.966	12.262	
WAM_CJ (CACE Comp 1) Year 5	0.375	0.021	167 (27)	3.692	162 (24)	3.357	3.527	
WAM_CJ (CACE Comp 2) Year 5	0.375	0.271	167 (27)	3.692	162 (24)	3.357	3.527	
WAM_CJ (Restricted to FSM Eligible) Year 5	0.103	0.325	42 (17)	3.39	49 (7)	4.215	3.835	

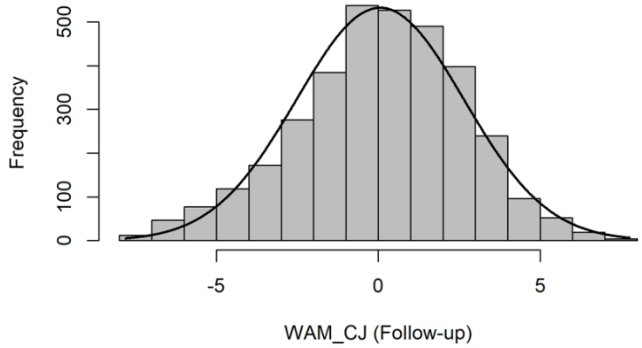
Appendix table C2.3: ICC estimation Year 2 Experiment

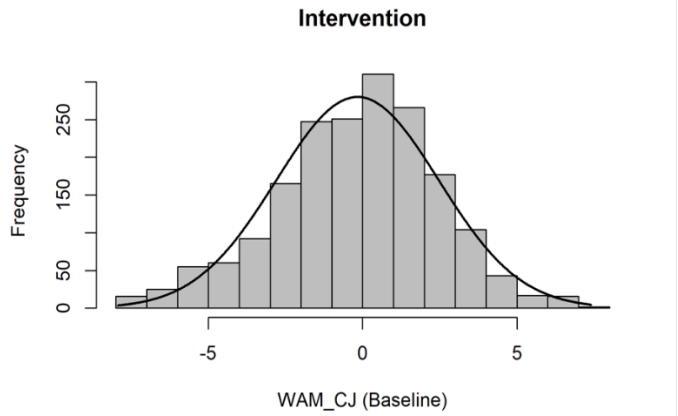
Outcome	Baseline	Follow-up
WAM_CJ	0.154	0.118
WAM_CR		0.238

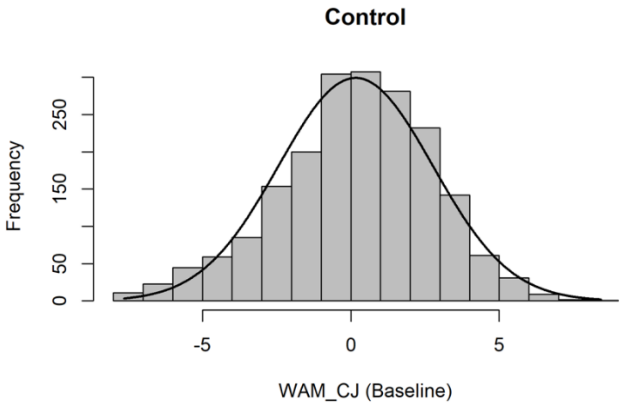
Appendix D1: Histograms and QQ plots for the Primary Outcome (Year 2 Experiment)

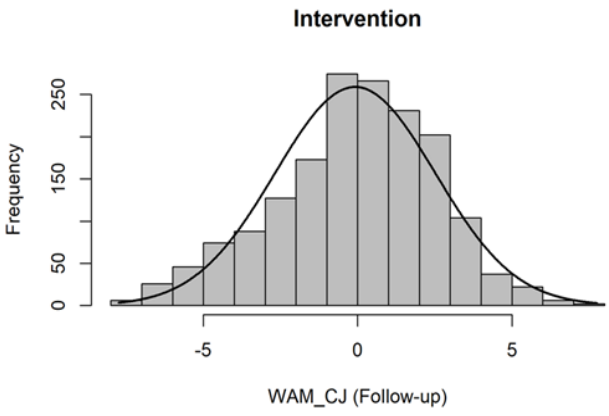
NB: including both intervention and control pupils.

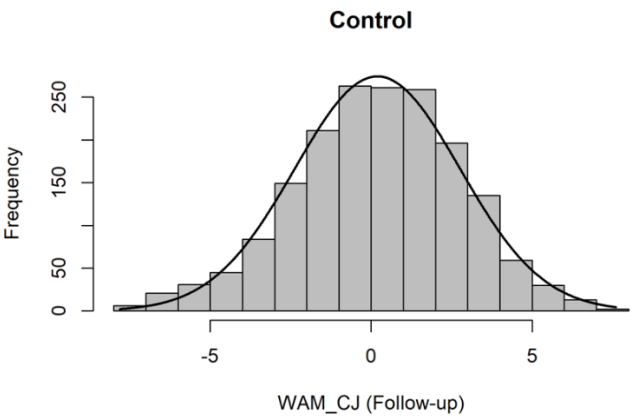


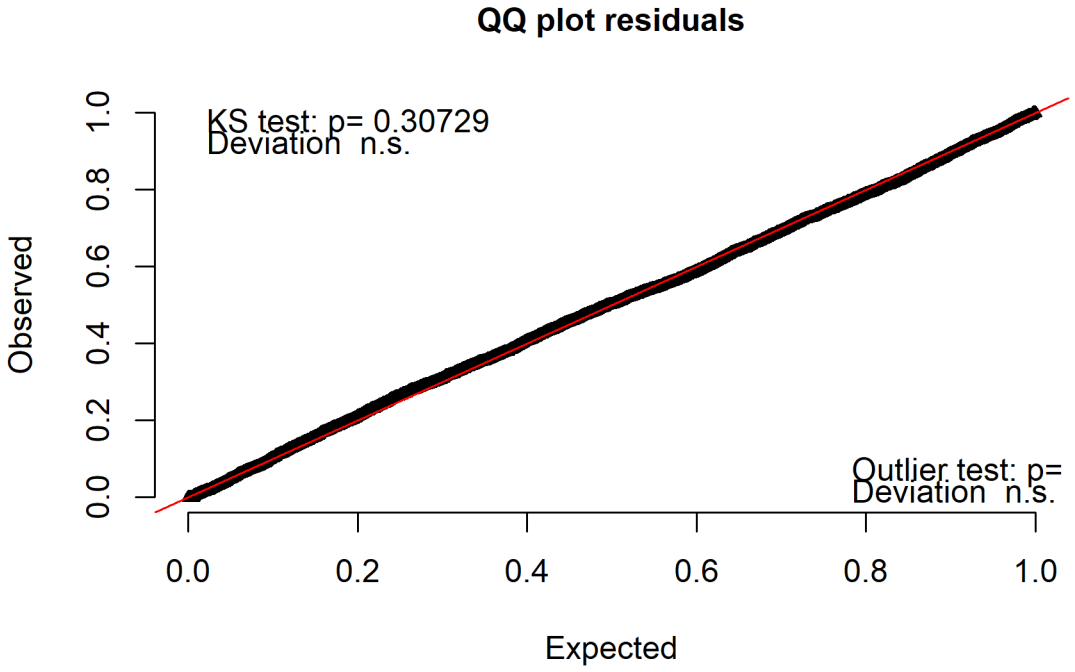




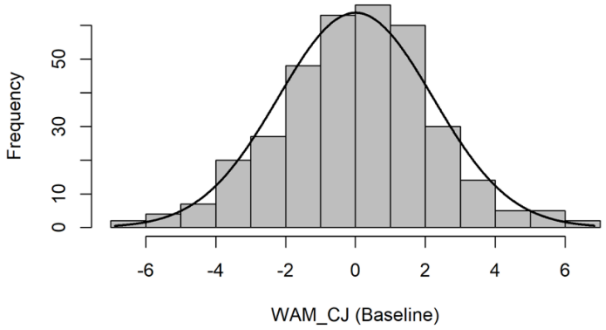


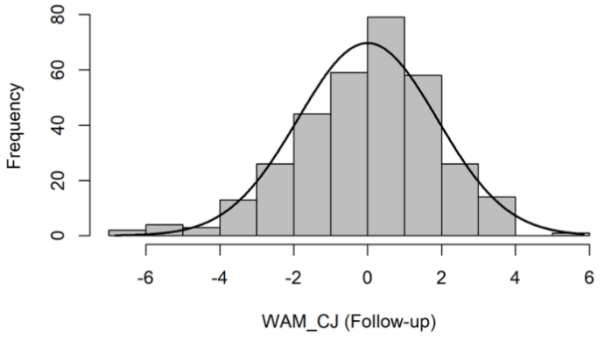


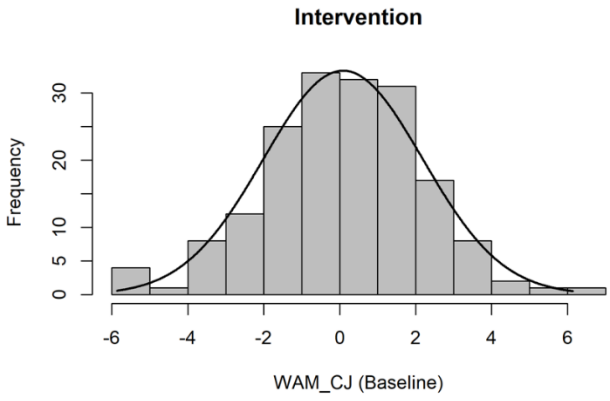


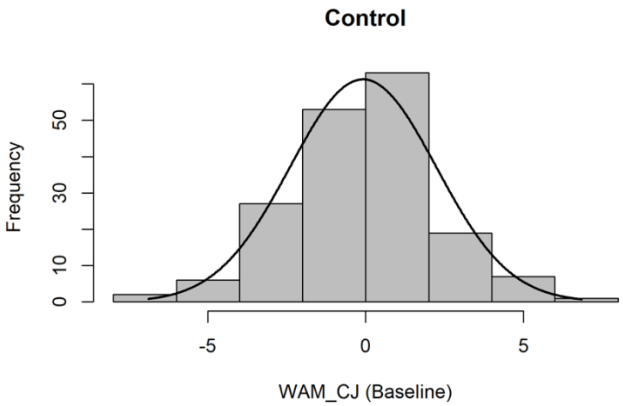

QQ plot residuals
KS test: p= 0.30729
Deviation n.s.
Outlier test: p=
Deviation n.s.
Observed
Expected
0.0
0.2
0.4
0.6
0.8
1.0

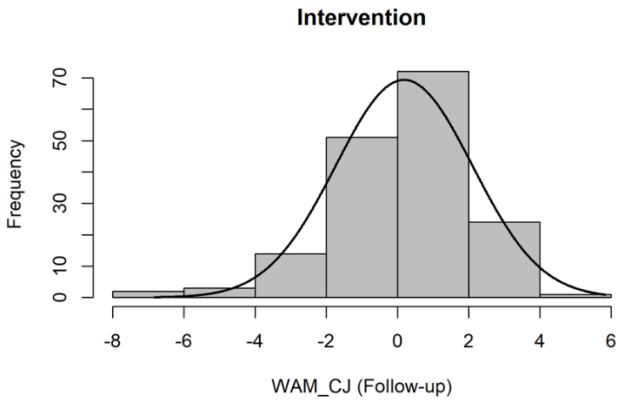
Appendix D2: Histograms and QQ plots for the Primary Outcome (Year 5 Experiment)

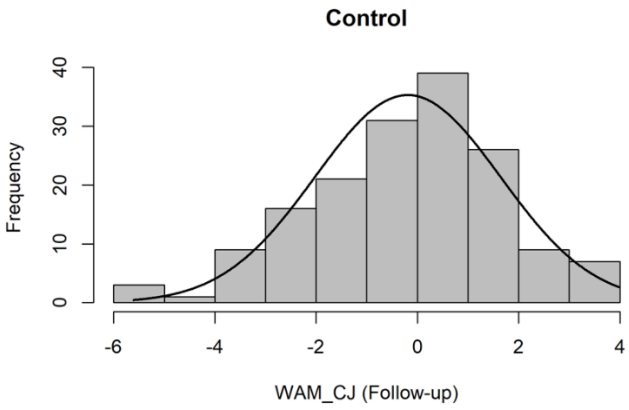


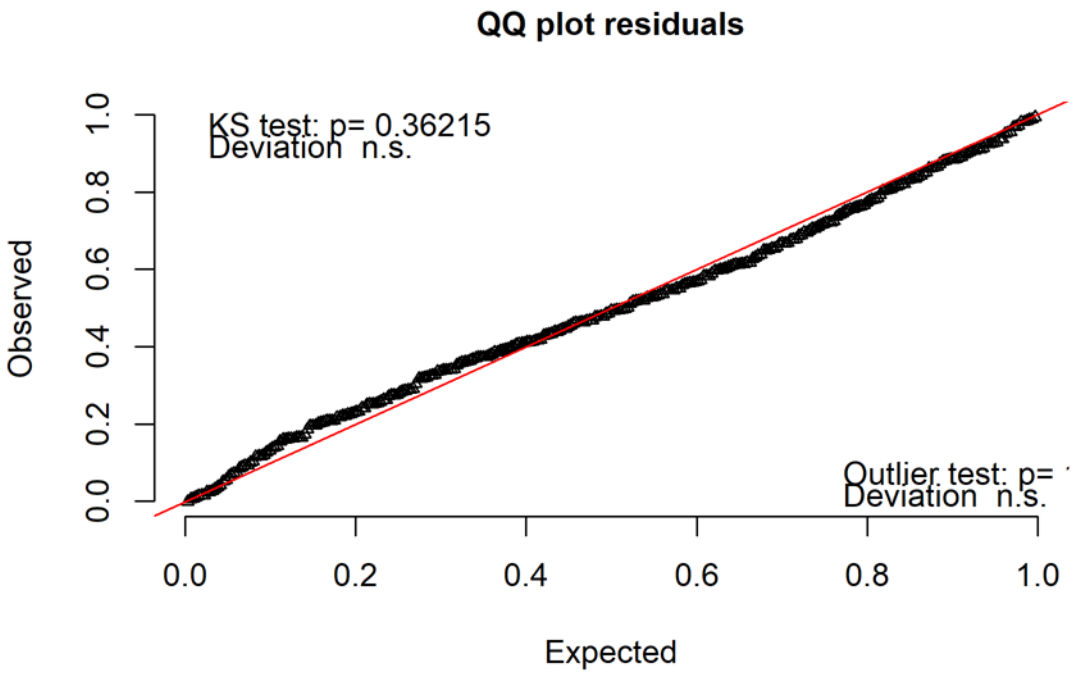




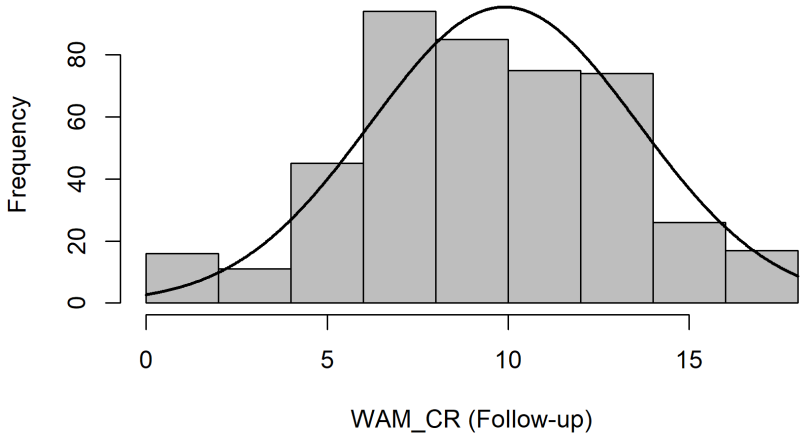


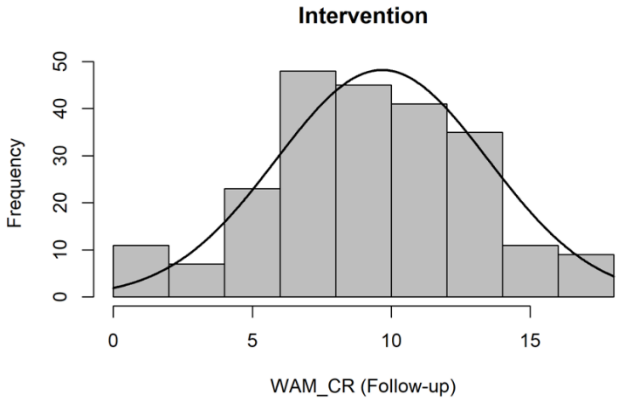


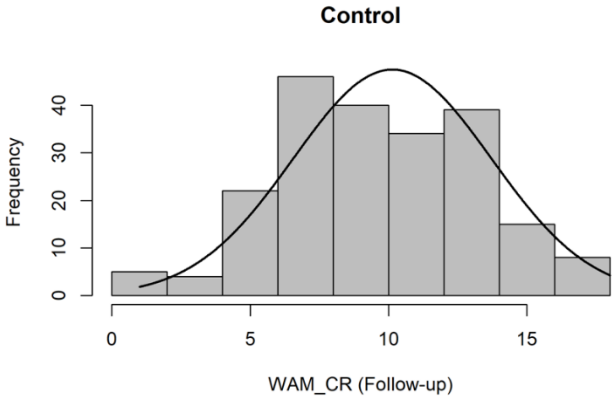


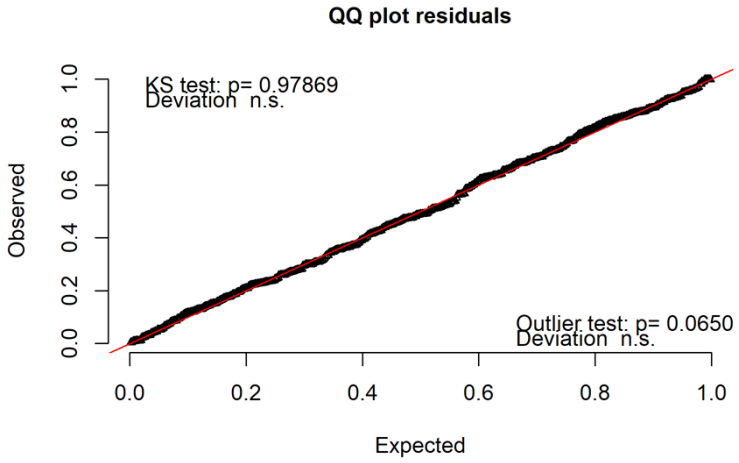

QQ plot residuals
KS test: p= 0.36215
Deviation n.s.
Outlier test: p=
Deviation n.s.
Observed
Expected
0.0
0.2
0.4
0.6
0.8
1.0

Appendix E1: Histograms and QQ plots for the First Secondary Outcome (Year 2 Experiment)

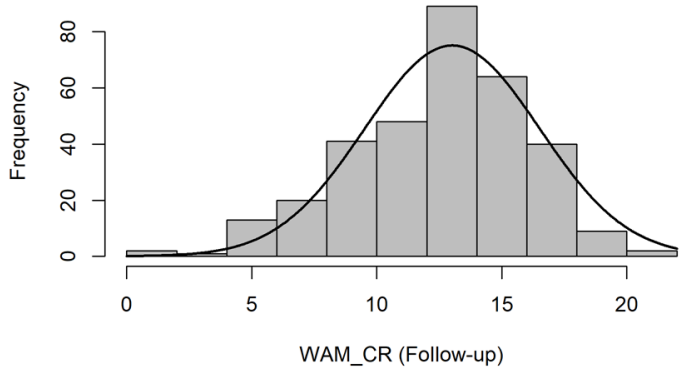


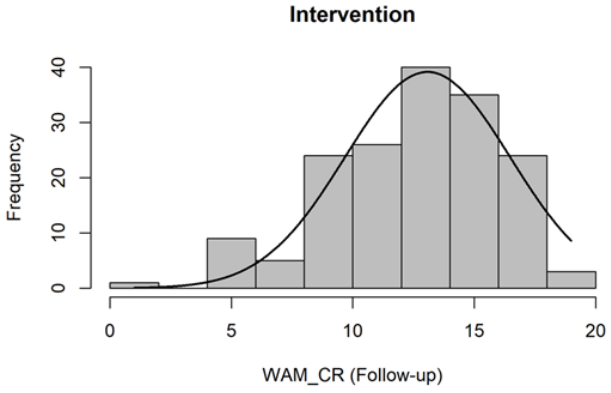


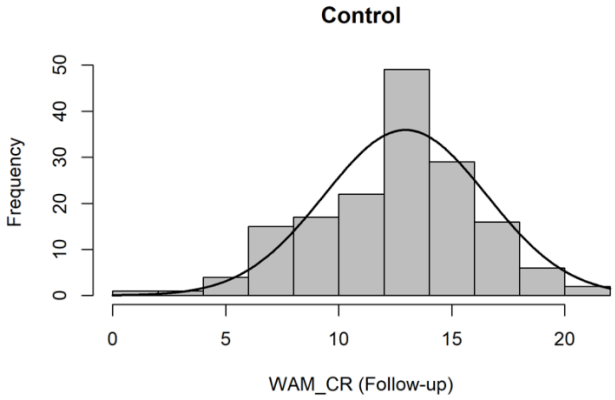


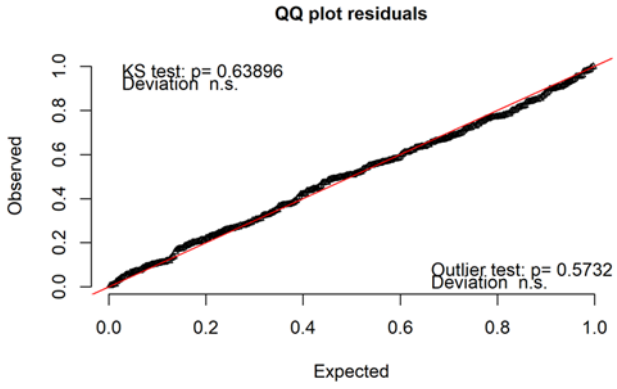


Appendix E2: Histograms and QQ plots for the First Secondary Outcome (Year 5 Experiment)

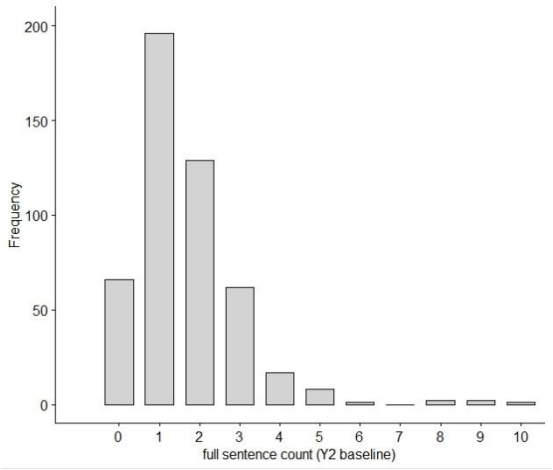


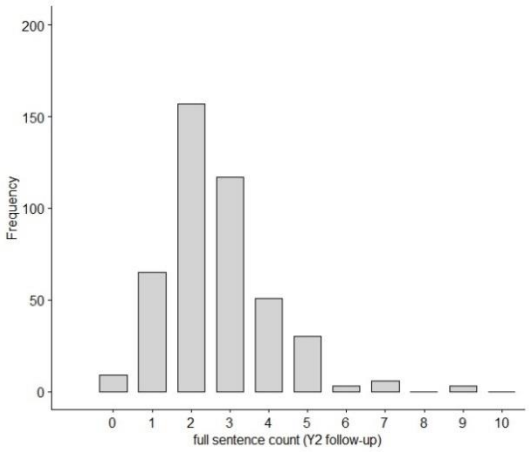


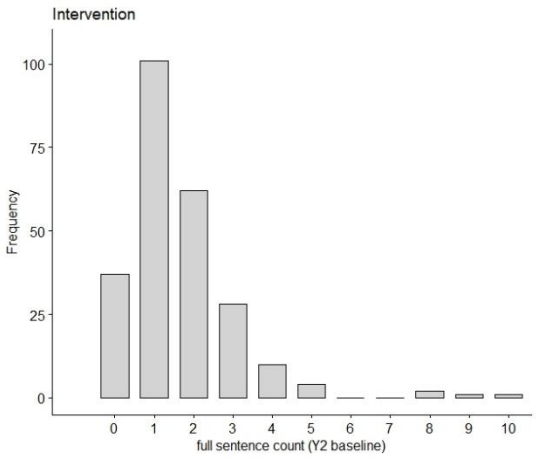


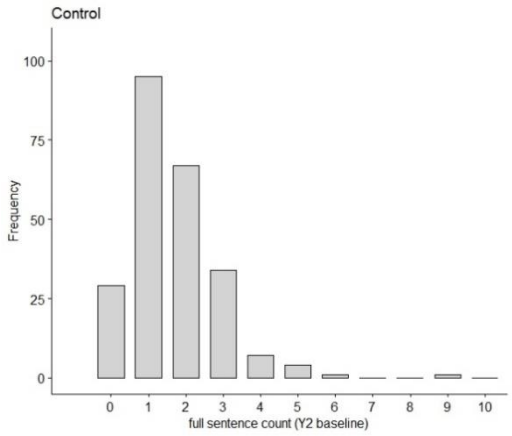


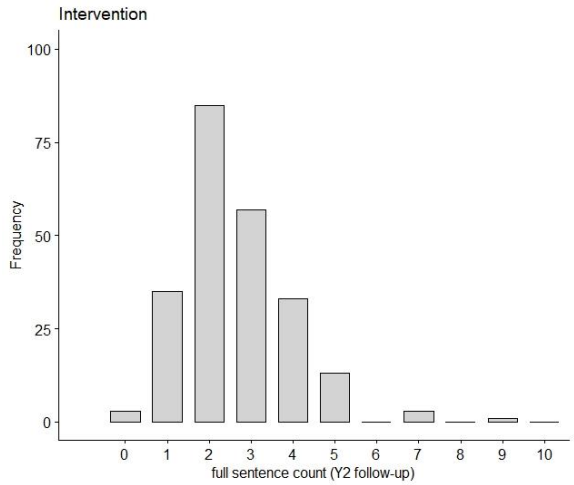
Appendix F1: Bar plots for the Second Secondary Outcome (Year 1 Experiment)

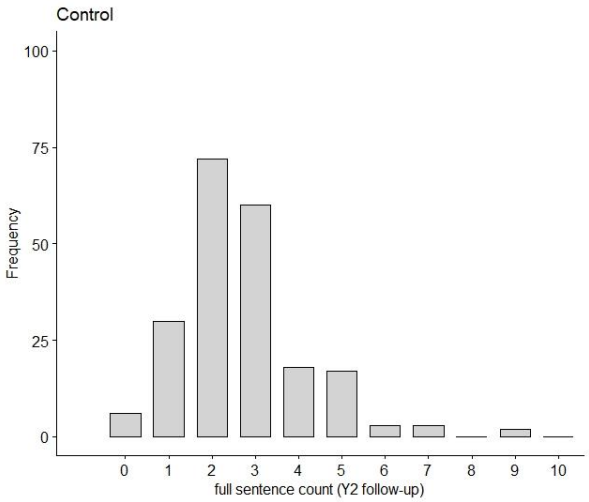




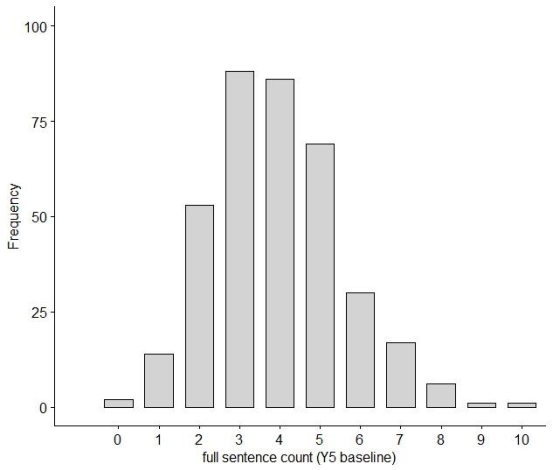


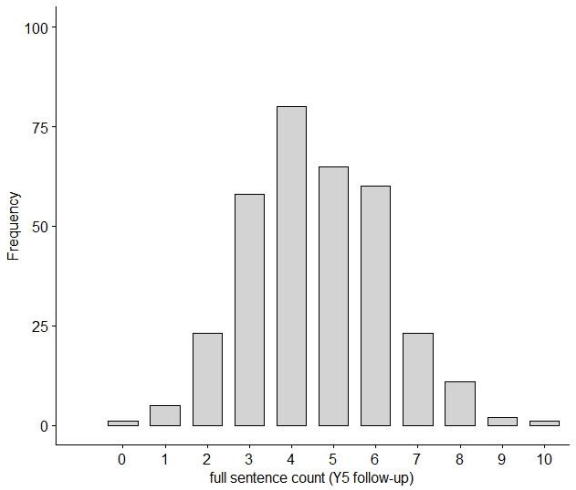


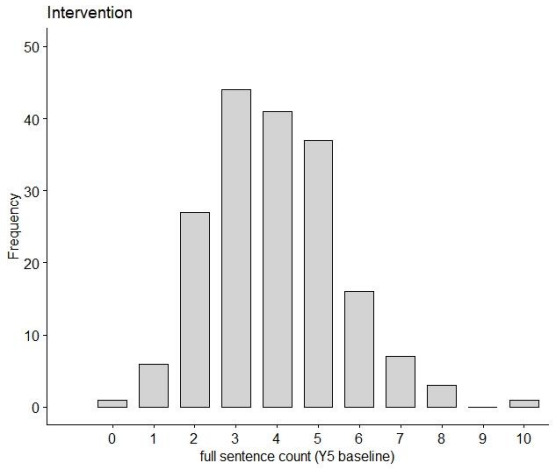


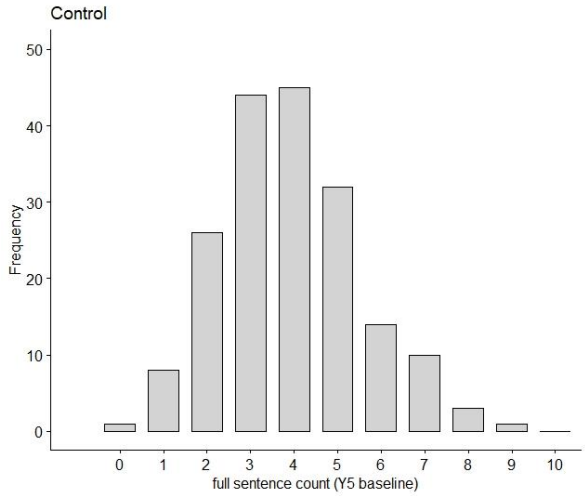


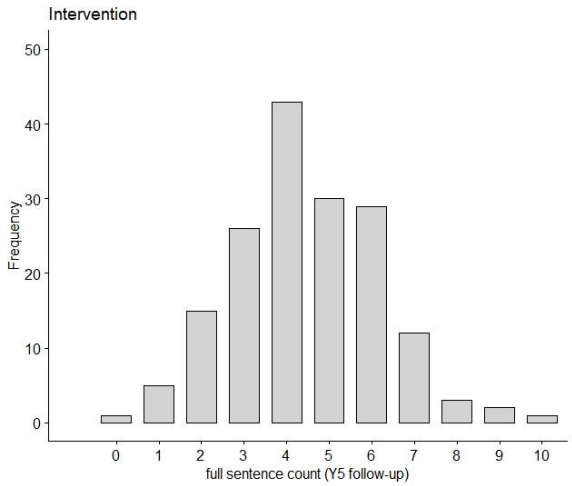
Appendix F2: Bar plots for the Second Secondary Outcome (Year 5 Experiment)

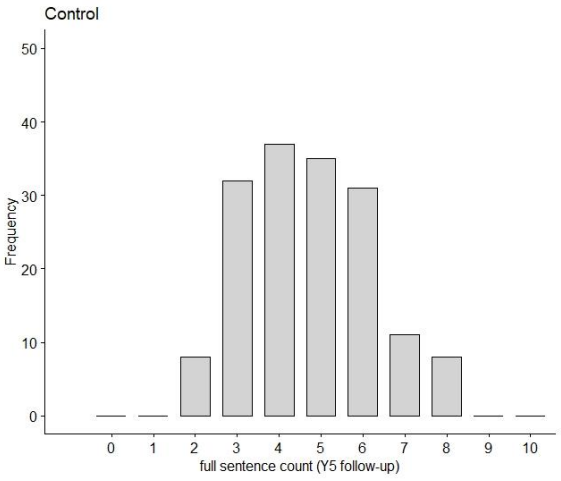












Appendix G: Privacy Notice

Privacy notice for Helping Handwriting Shine

Why are we collecting this data?

The National Foundation for Educational Research (NFER) are collecting personal data to enable the evaluation of the 'Helping Handwriting Shine' (HHS) intervention using a randomised controlled trial. The main aim of the HHS intervention is to increase children's capacity for automatic handwriting, which enables them to improve writing composition, and writing attainment overall. The trial aims to ascertain the impact of the intervention on pupil attainment in writing composition and handwriting speed.

Who is this research project sponsored and funded by?

The NFER is undertaking the independent evaluation of the HHS intervention, which is funded by the Education Endowment Foundation (EEF). NFER is joint data controller for this evaluation with the University of Leeds.

What is the legal basis for processing activities?

The legal basis for processing personal data is covered by:

> GDPR Article 6 (1) (f) which states that 'processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party except where such interest are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of the personal data'.

Our legitimate interest for processing your personal data is to administer the randomised controlled trial.

How will personal data be obtained?

Personal data will be collected by NFER directly from schools and through matching to the National Pupil Database. This will be augmented by assessment data provided both by the schools and through visits by NFER test administrators. NFER will also conduct case study interviews with teachers and pupils. University of Leeds will collect registers at training events that capture names and places of work.

What personal data is being collected by this project?

Personal data for the trial will include data about teachers and pupils from participating schools as described below:

NFER will collect data (name, job title and contact details) about a nominated named teacher and two staff members within a participating school so that we can liaise with them about the evaluation.

NFER will collect personal data about pupils. This includes pupil name, date of birth, unique pupil number (UPN), class name, school name for all pupils at Year 2 and for nominated pupils at Year 5. For nominated Year 5 pupils, Free School Meals status will also be collected. NFER will access pupil background data held on the National Pupil Database (NPD), DfE. The NPD data that we will request is pupil free schools meals (FSM) eligibility and gender.

NFER will match all of the above pupil data to pupil assessment data. This assessment data includes pupil results from two baseline tasks sat in Summer 2018; the Writing Assessment Measure and the Handwriting Speed test; and from the same two tests sat at the end of the trial in Summer 2019.

Schools will administer the baseline tasks in Summer 2018, and provide the pupil test papers to NFER. NFER's markers will mark the Handwriting Speed test and the Writing Assessment Measure. NFER will share the test results with schools.

NFER will conduct interviews at a selection of case study schools, with teachers and pupils. Interviewees will be anonymised in all reporting.

The above datasets will enable NFER to undertake primary and secondary outcomes analyses for the trial. NFER will share all of the above pupil data (pupil names, dates of birth, UPN matched to the NPD data described above and assessment results) with EEF's data archive partner. Anonymised data will also be stored in the UK Data Archive.

Who will the personal data be shared with?

Personal data will be uploaded to the No More Marking platform, upon which the Writing Assessment Measure will be marked.

NFER will scan the test papers and upload these, matched to each pupil's name, date of birth, school and UPN, into the No More Marking platform, where NFER markers will access and mark them. Markers will **not** have access to the personal data; test papers are presented anonymously.

Teachers' names and contact details will be shared between Leeds and NFER. This will be transferred via Secure File Transfer, an encrypted external portal accessed via a unique link sent to the intended respondent. The password for download will be provided in a separate e-mail.

Is personal data being transferred outside of the European Economic Area (EEA)?

No personal data is stored or transferred outside of the EEA.

How long will personal data be retained?

NFER, the University of Leeds and No More Marking will delete any personal data before three years from the completion of the project. (Note that retention of personal data is subject to agreement by the NPD team at the DfE).

NFER will send all the data it has collected to the Fisher Family Trust for archiving within three months of project completion, at which point EEF will take responsibility for Data Protection Compliance.

Can I stop my personal data being used?

NFER handles your personal data in accordance with the rights given to individuals under data protection legislation. If at any time you wish us to withdraw your data or correct errors in it, please contact handwritingRCT@nfer.ac.uk

In certain circumstances, data subjects have the right to restrict or object to processing, please contact our Compliance Officer. They also have the right to see information held about them. NFER will cooperate fully when a subject access request (SAR) is made.

Who can I contact about this project?

NFER is responsible for the day-to-day management of this project. Contact handwritingRCT@nfer.ac.uk with any queries.

In certain circumstances, data subjects have the right to restrict or to object to data processing. Please contact NFER's compliance officer in these circumstances. Individuals also have the right to see information held about them. You can make a subject access request by contacting NFER.

If you have a concern about the way this project processes personal data, we request that you raise your concern with us in the first instance (see the details above). Alternatively, you can contact the Information Commissioner's Office, the body responsible for enforcing data protection legislation in the UK, at https://ico.org.uk/concerns/.

Updates

We may need to update this privacy notice periodically so we recommend that you revisit this information from time to time. The date when this privacy notice was last updated is shown in the footer at the bottom of this document.

Appendix H: Data Sharing Agreement

Data Sharing Agreement

Project

EEFH: Helping Handwriting Shine Evaluation

What organisations are involved and what is their role?

Organisation	Role (e.g. data controller, joint data controller, data processor).
National Foundation for Educational Research (NFER)	Joint data controller
University of Leeds	Joint data controller
No More Marking	Data processor
Education Endowment Foundation (EEF)	Data controller (for archive stage – once the data have been submitted to the archive, after the trial has been completed)

What is the purpose of the data collection?

How does data sharing contribute to meeting this purpose?

The National Foundation for Educational Research (NFER) are collecting personal data to enable the evaluation of the 'Helping Handwriting Shine' (HHS) intervention using a randomised controlled trial. The main aim of the HHS intervention is to increase children's capacity for automatic handwriting, which enables them to improve writing composition, and writing attainment overall. The trial aims to ascertain the impact of the intervention on pupil attainment in writing composition and handwriting speed.

The legal basis for processing personal data is covered by GDPR Article 6 (1) (f) which states that;

> 'processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party except where such interest are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of the personal data'.

We have carried out a legitimate interest assessment which demonstrates that the evaluation fulfils one of NFER's core business purposes (undertaking research, evaluation and information activities) and is therefore in our legitimate interest, that processing personal information is necessary for the administration of the randomised controlled trial. We have considered and balanced any potential impact on the data subjects' rights and find that our activities will not do the data subject any unwarranted harm.

Description of data

What personal data will be collected?
What special data will be collected?
Which organisation will collect which data?

NFER will collect data (name, job title and contact details) about a nominated named teacher and two staff members within all participating schools so that NFER can liaise with them about the evaluation.

NFER will collect personal data about pupils. This includes pupil name, date of birth, UPN, class name, school name for all pupils at Year 2 and for nominated pupils at Year 5. For nominated Year 5 pupils, Free School Meals status will also be collected. NFER will access pupil background data held on the National Pupil Database (NPD), DfE. The NPD data that we will request includes pupil FSM eligibility and gender.

NFER will match all of the above pupil data to pupil assessment data. This assessment data includes pupil results from two baseline tasks sat in Summer 2018; the Writing Assessment Measure and the Handwriting Speed test; and from the same two tests sat at the end of the trial in Summer 2019.

NFER's markers will mark the Handwriting Speed test and the Writing Assessment Measure, matching outcomes with pupil data.

The above datasets will enable NFER to undertake primary and secondary outcomes analyses for the trial.

At the end of the trial NFER will share all of the data it collected with EEF's data archive processor, the Fischer Family Trust, for further analysis purposes. At this point EEF will become data controller. Anonymised data may also be submitted to the UK Data Archive at a later stage.

How will data be shared?

Data is shared using NFER's secure portal; a link to the upload portal will be provided to the data holder, with a password provided separately. The data holder will then upload the data where it will be encrypted while stored.

Data Protection responsibilities

	Relevant issues	Responsibilities
Fair and Lawful	What is the lawful basis for processing?	Legitimate Interests – see above
	Who will tell respondents that data sharing will take place?	NFER
	Who and how will the purposes of the processing be communicated to respondents?	NFER, via the Schools Information Sheet, Memorandum of Understanding and Privacy Notice.[51]
Processing for specified purposes	How will each organisation ensure that the data is only used for these purposes?	The data field and purpose for its collection is identified above and can only be collected and used for the identified purpose.
Adequate relevant and limited	How will each organisation ensure that they keep data collection to a minimum and share the minimum?	The data field and purpose for its collection/sharing is identified above and can only be collected and used for the identified purpose.
Keeping data accurate and up to date	Does the data need to be kept up to date, for further use, and how will this be done?	n/a
Retaining and deleting data	How long does each organisation need to keep the data and when will each delete the data?	Each organisation named above, excluding EEF, will delete any personal data after three years from the completion of the project. (Note that retention of personal data is subject to agreement by the NPD team at the DfE). NFER will send all the data to FFT archive within three months of the project completion. FFT will keep the data on behalf of EEF, at which point EEF take responsibility for Data Protection Compliance.
How will the data be kept secure	How will the data be exchanged between the organisations in a secure way? What other technical and organisational measures will be in place to keep data secure and to securely delete data in each organisation?	Data is shared using NFER's secure portal; a link to the upload portal will be provided to the data holder, with a password provided separately. The data holder will then upload the data. At the end of the project all appropriate data is placed into a folder which is permanently deleted from all locations (Project Closure).

Individual rights

How will data subjects be informed of their rights in relation to these processing activities
Who will be responsible for doing this?

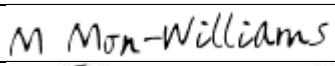
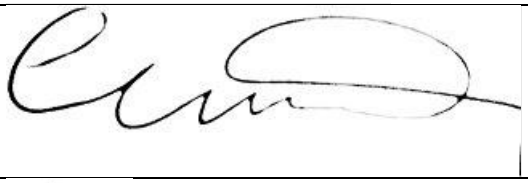
Data subjects are informed of their rights by the NFER, via the Schools Information Sheet, Memorandum of Understanding, Privacy Notice and a letter to parents of all children participating in the trial.

[51] The Privacy Notice is available online, with the link included on the Schools Information Sheet.

Transfer overseas – outside of the European Economic Area (EEA)
Personal data shall not be transferred to a country or territory outside the EEA unless that country or territory ensures an adequate level of protection for the rights and freedoms of data subjects in relation to the processing of personal data.

Will this data be transferred overseas outside the EEA	No
Who will be responsible for ensuring that data subjects know that this is the case	n/a
What measures will be put in place to ensure that the receiving country can be lawfully provided with that data	n/a

Agreed by

Organisation	Name	Signature*	Date
NFER	Maria Charles		20.4.18
University of Leeds	Mark Mon-Williams		4.4.18
No More Marking	Chris Wheadon		6.4.2018
EEF	Triin Edovald, Head of Evaluation		28/03/18

*signatures may be electronic

Appendix I: Memorandum of Understanding

Memorandum of Understanding

National Foundation for Educational Research
RPO, The Mere, Upton Park
Slough, Berkshire, SL1 2DQ
Telephone 01753 637205
Fax: 01753 790114

Email address: HandwritingRCT@nfer.ac.uk

RPO/EEFH/XXXX/2a

NFER No: «NFER_No»

Evaluation of Helping Handwriting Shine – Reply Form

If you would like to participate in the Evaluation of Helping Handwriting Shine, please read and sign the below reply form and Memorandum of Understanding (MOU), and return it to NFER using the pre-paid envelope or scan and email to HandwritingRCT@nfer.ac.uk

	Contact's Details	Please amend if necessary
School Name:		
Headteacher:		
Tel. No:		
Fax No:		
Email address:		

My school **will** take part in this evaluation and agrees to the conditions stated in the Memorandum of Understanding (MOU).

Headteacher / SMT signature…….....……

Name of nominated Helping Handwriting Shine contact in the school:
Mr/Mrs/Miss/Ms/Dr: .....
Job title: …….....
Contact phone number: ……..……
Contact email address: ……..……
Best time to contact you: .....

How many students are currently in Year 1 and Year 4: ☐ ☐

How many classes are currently in Year 1 and Year 4: ☐ ☐

Once we have received your reply form we will send you a confirmation email confirming receipt and outlining the next steps.

NFER No: «NFER_No»

School Name
School Address
Schools Address
School Postcode

RCT Evaluation of Helping Handwriting Shine programme for Schools Memorandum of Understanding

The following outlines our expectations from schools and teachers taking part in the evaluation. Please read the following statements and sign the reply form provided to confirm that you have read the document, please also sign and keep this copy for your reference.

Our overall expectations of the school;

- The school must allocate a named contact to the project to work with NFER and the Helping Handwriting Shine Team (guidance can be provided on who this should be). They should have sufficient capacity to be able to respond promptly to requests and facilitate requirements as appropriate. If they leave the school or are no longer able to meet the requirements for the role, NFER must be promptly informed of this, and details of a replacement contact.
- All data required by the evaluation team in relation to the project must be provided in a timely fashion.
- For the purpose of research, the responses will be linked with information about your pupils from the National Pupil Database (held by the Department for Education), other official records, and shared with NFER, the University of Leeds, the Department for Education, EEF's data contractor FFT Education and in an anonymised form to the UK Data Archive and for research purposes. Your pupil's data will be treated with the strictest confidence. Neither we, nor the named parties, will use pupil names or the name of the school in any report arising from the research.

Specific expectations of all schools

- Schools will undertake the baseline testing for Year 2 (current Y1) in Summer 2018 and for Year 5 in Autumn 2018.
- Schools will allow an NFER Test Administrator to attend their schools in June/July 2019 to administer the end point testing.

Schools allocated to the Intervention group during the 2018/19 Academic Year

- The school will pay £500 in order to access the intervention and training.
- A minimum of two staff identified to take part in the project will attend one half-day training session during October 2018.
- The trained staff will deliver three thirty-minute intervention sessions per week for eight weeks between November 2018 and February 2019, for all Year 2s and the selected Year 5 pupils.
- Trained staff will complete weekly logs to evidence their practice in the intervention.
- School will enable the selected staff members to have sufficient time to undertake the programme

Expectations of Schools allocated to the Control group during the 2018/19 Academic Year

- Trained staff will complete logs to evidence their Business as Usual practice.

Timetable of Activities for Schools

Date	Activity
March – May 2018 **All schools**	**Sign-up to take part in the trial** • Return reply form and signed Memorandum of Understanding to NFER in the pre-paid envelope • Schools take a copy of the MoU to keep for their records • Schools provide NFER with class list of all current Y1 pupils (Year 2 2018/19). • Schools write to parents to let them know the school is participating in the intervention

June / July 2018
All schools
	Pupil baseline writing assessments for all pupils in Y1
	• A test consisting of two tasks:
	  ○ Writing Assessment Measure – 20 minutes
	  ○ Handwriting speed test – 3 minutes
	• Schools will receive prepopulated papers via secure courier from NFER. Once the test is completed schools can arrange for pre-paid courier to collect completed test papers and return to NFER, who will organise for them to be marked through standard mark schemes and also a comparative judgement approach.
	• Proforma about existing and intended handwriting provision for Year 2 and target Year 5 pupils to be completed.
	Only when a school has completed all the above steps have they completed the sign up process. At this point schools will receive a confirmation email from NFER confirming that their school will go forward to randomisation.

July 2018 **All schools**	**Schools randomly allocated** to the intervention or control group. • Schools will receive an email during the week commencing the 16th of July confirming which group they have been allocated to and what the next steps will be. • For schools allocated to the intervention group they will be asked to provide the pupil lists for the identified Y4 (Year 5 2018/19) pupils.
September 2018 **Intervention schools**	**Schools provide final lists of identified Year 5 pupils** • All identified Year 5 pupils then complete the baseline test and tests are returned to NFER. • Year 5 pupils randomised in each school with half to receive the intervention and half to continue as normal.
October 2018 Intervention schools	**Schools book and attend training sessions** • Schools book onto training sessions, paying £500 per school. • Sessions will be run at local centres across Leeds, Bradford and two locations in the North East, to be confirmed. • A minimum of two members of staff must attend, who may be teachers, senior leadership, SENCOs, teaching assistant, occupational therapists, etc. • Schools are asked to cover travel costs and staff cover costs for these events. • These training sessions will be observed by the NFER evaluation team.
Nov 2018 – Feb 2019 **Intervention schools**	**Schools deliver the programme** • Programme delivery to Year 2 (whole class/es) and selected Year 5 • Staff complete online logs which outline what sessions have taken place and which pupils attended. • Further training and support provided. • Observations in some schools.
Spring term 2019 **Control schools** **Intervention schools**	**Control schools complete** 'Business as Usual' logs **Intervention schools embed approaches** • Some intervention schools take part in interviews and observations • Staff complete online logs • Complete follow up tasks to embed approaches
Summer 2019 **ALL Schools**	**Pupil follow-up tests** • All Year 2 – both control and intervention • Selected Year 5 pupils – intervention schools • Administered by NFER test administrators

Please read the following statements and sign at the bottom if you agree with the statements:

I confirm that I have read and understand the information provided about the project and I have passed a copy of the Memorandum of Understanding and School Information Sheet to my designated named contact. I have had the opportunity to ask questions, and have had these answered satisfactorily.

This is a project that aims to evaluate the impact of the Helping Handwriting Shine Programme; I understand that my school will be randomly assigned to either an intervention or a control group. The project is to be delivered in academic year 2018-19.

I understand that my school's participation is voluntary and that I am free to withdraw my school at any time, however I will let NFER know if I choose to withdraw from the trial.

I agree to facilitate the activities involved in the evaluation project as described above and in the Project Information Sheet.

I know whom I can contact if I have any concerns or complaints about the study.

I agree that my school will take part in the above study.

Signed……

Print……

Position……..

Date……

Appendix J1: School Information Sheet

School Information Sheet

RCT Evaluation of Helping Handwriting Shine

Information for schools

What is Helping Handwriting Shine?

Helping Handwriting Shine focuses on improving crucial elements of fine motor skills in order to develop children's handwriting fluency. In so doing, it aims to increase children's capacity to develop writing composition, and to improve children's writing attainment overall. Members of school staff are trained to deliver the programme to all Year 2 and selected Year 5 pupils; they then deliver the programme to pupils, which consist of three 30-minute sessions per week for eight weeks, over Autumn 2018 - Spring 2019. Sessions include a selection of engaging activities and approaches to support children's handwriting (including body-position, pattern formation, stylus grip, and breathing). Training materials provided by the programme offer guidance on how sessions can be adapted to support pupil interest and motivation, and to mirror curriculum themes.

What are the trial aims?

The programme is being evaluated through a randomised controlled trial (RCT) to explore the impact on children's overall writing, handwriting speed, and writing composition. The evaluation will also investigate how the sessions are delivered, and teachers' and children's views on the programme and the support provided.

Who is conducting the trial?

The Education Trials Unit at the National Foundation for Educational Research (NFER) is conducting the trial. The University of Leeds is providing the training and support to schools for the Helping Handwriting Shine intervention. The Education Endowment Foundation (EEF) is funding the trial. The NFER has subcontracted No More Marking to provide test outcomes.

What will the trial involve for schools?

In Spring/Summer 2018, all schools will need to provide pupil lists/IDs for all current Year 1 pupils; Year 1 pupils will then need to complete baseline tests, and the school completes a short pro-forma. Schools will then be **randomly allocated** to either the intervention group (who receive the programme) or the control group (who do not receive the programme).

In Autumn 2018, Intervention schools select Year 5 pupils to take part in the intervention. Between 4 and 16 pupils should be selected depending on the school's size and needs. Schools provide pupil lists/IDs for selected pupils, who then complete the same baseline test as the Year 2 classes; they will then be randomised so that half of the group receive the intervention and half do not.

Intervention schools will then book a minimum of **two** identified members of staff onto a half-day training session, which will be held locally to the school, paying a discounted rate of £500 per school to receive the training and materials. Schools will then commence the 8-week programme, split into two four week sessions either side of the Christmas holidays. Intervention schools will need to complete a weekly record about the sessions; and be willing to take part in observations and interviews if invited. On completion of

the eight week programme there will be a number of tasks and activities for pupils before follow-up tests are carried out in Summer 2019.

Control schools will not receive the handwriting intervention, but will need to complete records about any handwriting support they provide in their schools during the trial period. Control schools may also be invited to take part in interviews and observations, and will be required to complete the follow-up test in June/July 2019. Control schools will receive a contribution of £500 on completion of the follow-up tests.

In Summer 2019, NFER test administrators will then support **all schools** to complete follow-up tests with all pupils in the trial. This includes all selected Year 5 pupils selected to take part within the intervention schools, whether or not they have received the intervention.

Which schools, staff and pupils will be involved?

The trial will involve around 100 primary schools across Leeds, Bradford and North East England. The trial is for all Year 2 pupils (with the intervention delivered to all Year 2 whole classes in the school), and also for small groups of targeted Year 5 pupils in intervention schools in need of handwriting support, where slow and effortful handwriting is interfering with their ability to communicate with others.

The school must identify a minimum of two members of staff to be trained and deliver the intervention. The members of staff need not necessarily be teachers; where more appropriate they may be, for example, members of senior leadership, SENCOs, teaching assistants or occupational therapists. This may be particularly important for Year 5, where delivery of the intervention may happen separately to normal teaching.

Each school will also need to provide a named school contact for the duration of the trial. This person will act as the single point of contact between the school and the developers and evaluators of the trial and could be a class teacher, teaching assistant, SENCO or a member of the leadership team. The named contact will ensure the return of baseline assessments, ensure the number of sessions are delivered in the way prescribed and that the logs are completed and returned for evaluation. The contact will also be responsible for liaising with NFER in Summer 2019 for the administration of the follow-up tests.

Timetable of Activities for Schools

Date	Activity
March - May 2018 **All schools**	**Sign-up to take part in the trial** • Return reply form and signed Memorandum of Understanding to NFER in the pre-paid envelope • Schools take a copy of the MoU to keep for their records • Schools provide NFER with class list of all current Y1 pupils (Year 2 2018/19). • Schools write to parents to let them know the school is participating in the intervention

June / July 2018
All schools
	Pupil baseline writing assessments for all pupils in Y1
	• A test consisting of two tasks:
	  ○ Writing Assessment Measure – 20 minutes
	  ○ Handwriting speed test – 3 minutes
	• Schools will receive prepopulated papers via secure courier from NFER. Once the test is completed schools can arrange for pre-paid courier to collect completed test papers and return to NFER, who will organise for them to be marked through standard mark schemes and also a comparative judgement approach.
	• Proforma about existing and intended handwriting provision for Year 2 and target Year 5 pupils to be completed.
	Only when a school has completed all the above steps have they completed the sign up process. At this point schools will receive a confirmation email from NFER confirming that their school will go forward to randomisation.

July 2018 **All schools**	**Schools randomly allocated** to the intervention or control group. • Schools will receive an email during the week commencing the 16th of July confirming which group they have been allocated to and what the next steps will be. • For schools allocated to the intervention group they will be asked to provide the pupil lists for the identified Y4 (Year 5 2018/19) pupils.
September 2018 **Intervention schools**	**Schools provide final lists of identified Year 5 pupils** • All identified Year 5 pupils then complete the baseline test and tests are returned to NFER. • Year 5 pupils randomised in each school with half to receive the intervention and half to continue as normal.
October 2018 **Intervention schools**	**Schools book and attend training sessions** • Schools book onto training sessions, paying £500 per school. • Sessions will be run at local centres across Leeds, Bradford and two locations in the North East, to be confirmed. • A minimum of two members of staff must attend, who may be teachers, senior leadership, SENCOs, teaching assistant, occupational therapists, etc. • Schools are asked to cover travel costs and staff cover costs for these events. • These training sessions will be observed by the NFER evaluation team.
Nov 2018 – Feb 2019 **Intervention schools**	**Schools deliver the programme** • Programme delivery to Year 2 (whole class/es) and selected Year 5 • Staff complete online logs which outline what sessions have taken place and which pupils attended. • Further training and support provided. • Observations in some schools.

Spring term 2019 **Control schools** **Intervention schools**	**Control schools complete** 'Business as Usual' logs **Intervention schools embed approaches** • Some intervention schools take part in interviews and observations • Staff complete online logs • Complete follow up tasks to embed approaches
Summer 2019 **ALL Schools**	**Pupil follow-up tests** • All Year 2 – both control and intervention • Selected Year 5 pupils – intervention schools • Administered by NFER test administrators

How will schools benefit from taking part?

All schools in the trial will contribute to the evidence-base on what works in supporting handwriting and writing in primary schools. Schools allocated to the intervention group will have the opportunity to receive discounted training and support on the Helping Handwriting Shine intervention. Schools will also receive feedback on relative pupil performance in both sets of tests.

Do schools have to take part?

Helping Handwriting Shine is available only as part of the trial during 2018/19. Schools only have to take part if they wish to do so. Pupils can opt out of taking part at any time. However, all data is important to the trial, and the NFER and the University of Leeds really appreciate schools' support in providing pupil tests data for the trial.

How will NFER and partners use and protect the data collected?

For information on how the parties involved in this research will gather, use and protect data, please refer to the Privacy Notice for the Evaluation of Helping Handwriting Shine, available at https://www.nfer.ac.uk/pdf/EEFH_Privacy_Statement.pdf

How will the findings be used?

The findings from the project will be freely available on NFER's and EEF's website. They will be used to inform the education sector about improving writing in primary schools.

Who can I contact for more information?

Who to contact		Telephone	Email
Gemma Stone/Pippa Lord, NFER	For queries about the trial	01904 567647	g.stone@nfer.ac.uk
Jishi Jose, NFER	For queries about signing up and providing data	01753 637205	HandwritingRCT@nfer.ac.uk
University of Leeds Project Team	For queries about the Helping Handwriting Shine programme		handwriting@leeds.ac.uk

Appendix J2: Dos and Don'ts poster

Helping Handwriting Shine

This research project involves Year 2 and Year 5 pupils. In Year 5, some pupils receive the programme, and some pupils do not receive the programme but are still part of the research. To avoid *contamination* between these two groups, please take note of these guidelines and follow them to the best of your ability.

- **Don't** use any of the content, terminology or techniques from the programme when talking to or teaching any pupils who are not receiving the programme.
- **Don't** talk to other staff, who are not delivering the programme, about the programme's content, terminology or techniques.
- **Don't** display any pupil's work if it shows any work resulting from the programme.
- **Don't** put any of the pupil's work on shared or public online areas if it shows any work resulting from the programme.
- **Don't** talk about the programme with pupils receiving the programme in front of pupils who are not receiving the programme.
- **Do** take out the pupils receiving the programme out of normal classes – **don't** teach them in the same room as the rest of the class.
- **Do** make sure that pupil's work resulting from the programme is kept secure and separate to other pupils' work.
- **Do** make sure that the handbook and any other programme materials are kept secure, ideally in a locked cabinet or drawer.
- **Do** remember to fill in the logs for both Year 2 and Year 5, and make a note of any potential contamination.

*Email **handwriting@nfer.ac.uk** if you have any queries.*

Appendix K: Year 5 pupil selection guidance

Helping Handwriting Shine: Selecting Year 5 pupils

School Name:

Number of pupils:

This guidance will help the school's Year 5 teacher select the pupils who will be part of the trial. As the school's Year 5 teacher, we believe you are best able to select the pupils. You may wish to confer with the class' Year 4 teacher, or another staff member who knows the pupils well.

Please read the following information carefully. If you need any guidance in selecting pupils, or have any questions, please contact us at **handwritingRCT@nfer.ac.uk**

There are lots of children with poor manual dexterity who struggle with the complexities of holding a pen or pencil and producing legible handwriting. Perhaps this is not surprising – handwriting is an extremely difficult task that requires precisely controlling the forces applied through the fingertips whilst generating hand movements that need to be coordinated across space and time. The fact that most adults can write without difficulty is a testament to the incredible ability of humans to learn complex motor skills. This can mask the barriers that many children struggle to overcome when acquiring this skill.

Helping Handwriting Shine is designed for pupils with such motor problems and who consistently demonstrate illegible and/or slow and/or effortful handwriting. These pupils can be contrasted with other pupils who have the ability to produce legible handwriting but who show messy handwriting because of a lack of care or a strategic decision to focus their efforts on other aspects of the task beyond producing neat handwriting (but still produce legible scripts).

Teachers are very good at distinguishing between these two groups and are well positioned to identify children with motor deficits who struggle with the **mechanics** of handwriting per se (and thus spend a long time in writing exercises, overly focus their efforts on the writing process and/or produce illegible script).

It is these children who will show the greatest benefit from the handwriting intervention programme. Identify them using the following guide.

1. Select pupils that have **slow** or **effortful handwriting**.
2. Avoid selecting children who have messy handwriting for behavioural, or other, reasons.
3. If this is more than your target number of pupils, try to 'rank order' the pupils to identify which could benefit most from the intervention.
4. If you need to include more pupils to meet your target number, consider any pupils that may have slightly faster or less effortful handwriting, but cannot read their own handwriting afterwards.

Appendix L: Parent opt out letter

Parent opt out letter

RPO/EEFH/55015-16-21/2Opt

Dear Parent / Guardian

We are writing to let you know that your child's school has agreed to be part of a research trial, investigating the efficacy of the Helping Handwriting Shine programme. Helping Handwriting Shine focuses on improving crucial elements of fine motor skills in order to develop children's handwriting fluency. It also aims to increase children's capacity to develop writing composition, and writing attainment overall.

The National Foundation for Educational Research (NFER) have been asked to evaluate the programme by the Education Endowment Foundation (EEF) through a randomised controlled trial. The University of Leeds is providing the training, materials and support to schools for the trial. The programme is being evaluated to explore the impact on children's overall writing, handwriting speed, and writing composition. The evaluation will also investigate how the sessions are delivered, and teachers' and children's views on the programme and the support provided.

If you are happy for your child's data to be used for this research, **you do not need to return the reply slip**. If you would prefer your son/daughter's data not to be shared, stored and used for research purposes, please complete the form below. Please inform their teacher if you would like to withdraw your son/daughter's data from the evaluation at any stage.

For further information about this trial please refer to https://www.nfer.ac.uk/for-schools/participate-in-research/rct-evaluation-of-helping-handwriting-shine/. NFER have robust procedures in place to make sure that we comply with the increased governance and accountability requirements of GDPR. For further information on how this research will gather, use and protect data, please see the project's Privacy Notice at https://www.nfer.ac.uk/pdf/EEFH_Privacy_Statement.pdf. If you have any queries please contact us via email at HandwritingRCT@nfer.ac.uk.

Yours sincerely

Kathryn Hurd
Head of Survey Operations
National Foundation for Educational Research

Evaluation of Helping Handwriting Shine
OPT-OUT SLIP – you only need to complete this form if you do NOT wish your child's data to be stored and used for research purposes.

I DO **NOT** give permission for data about my child that is collected as part of the Evaluation of Helping Handwriting Shine project to be shared, stored or used for research purposes.

Your child's name……Child's class:… …….

Name of School……...

Your full name……..

Your telephone number (optional)....…….

Your signature…… Date……....

Appendix M: Writing Assessment Measure marking scheme

Writing Assessment Measure marking scheme (Dunsmuir, Kyriacou, Batuwitage, Hinson, Ingram and O'Sullivan, 2013).

Writing Assessment Measure (WAM)

TIME GUIDELINE: *Prompt 1:* 15 minutes *Prompt 2:* 15 minutes
DISCONTINUE RULE: Stop the child after 15 minutes of writing

Elements and Criteria	Circle Score
Handwriting	
• Writing is consistent, fluent and cursive.	4
• Clear, neat and legible and may show evidence of joining handwriting	3
• Handwriting may vary in shape and size and is beginning to develop consistency.	2
• Handwriting is indecipherable or difficult to read.	1
Spelling	
• Evidence of correct spelling of complex words containing prefixes/suffixes or irregular words e.g. souvenir, destruction, and conscious.	4
Attempts to spell some complex or polysyllabic words using visual or phonetic strategies, e.g 'safariye' for safari, 'adventerous' for adventurous.	3
• Spells the majority of high frequency common words correctly e.g. inside, because, while.	
• Spells some common monosyllabic words correctly (e.g. mum, cat, bird). Uses phonic strategies to attempt to spell high frequency common words e.g. 'grat' for great, 'fhun' for fun.	2 1
Punctuation	
• Uses a range of punctuation to clarify structure and create effect (e.g. speech marks, dashes, brackets, apostrophes, commas to demarcate sentences).	4
• Secure use of full stops and capital letters. Uses punctuation in addition to capital letters and full stops, the majority are used correctly (e.g. question marks, exclamations marks, commas in lists).	3
• Evidence of accurate use of capital letters and full stops, however few there are. (e.g. Sentence finishes with a full stop and next sentence begins with a capital letter)	2
• Shows awareness of how full stops are used in writing.	1
Sentence Structure and Grammar	
• Secure control of complex sentences. Understands how clauses can be manipulated for effect. Able to use conditional and passive voice (e.g. having watched him eat a dog biscuit, she felt sick)	4
• Beginning to write extended sentences including subordinators (e.g. if, so, while, when, after). The basic grammatical structure of sentences usually correct (e.g. usually consistent and correct use of tenses and nouns and verbs agree).	3
• Beginning to use other conjunctions to create compound sentences (e.g. because, but, so, then) and may be using multiple clauses (still mixing up tenses).	2
• Writes simple sentences which include the conjunction 'and'.	1
Vocabulary	
• Demonstrates use of well-chosen vivid & powerful vocabulary to create effect (e.g. verbs, adjectives, adverbs)	4
• Varied use of adjectives, verbs and specific nouns (e.g. delicious for nice/sauntered for went/poodle for dog)	3
• Some selection of interesting and varied verbs e.g. jumped, compare, guess	2
• Uses simple vocabulary, appropriate to content. Writing is composed of simple nouns and verbs e.g. look, went, go, play, see	1
Organisation and Overall Structure	
• Paragraphs are well organised, based on themes and provides a cohesive text for the reader (e.g. paragraphs, subheadings, logically organised events).	4
• Uses paragraphs to organise writing, showing an identifiable structure. May be short sections.	3
• Themes are expanded upon and linked together in a series of sentences.	2
• Communicates meaning but may 'flit' from idea to idea and any themes that are expanded are done so in one sentence.	1
Ideas	
• Ideas are creative and interesting in a way that engages the reader. Uses a range of strategies and techniques such as asides, comment, observation, anticipation, suspense, tension.	4
• Ideas are imaginative and varied evidence of descriptive detail about characters, settings, feelings, emotions & actions.	3
• Ideas are developed to by adding detail (e.g. is beginning to provide additional information or description beyond a simple list).	2
• Produces short sections of ideas which may be repetitive and limited in nature.	1
Total score	

Appendix N: Amended

Writing Assessment Measure marking scheme

Spelling	Mark
• Evidence of correct spelling of complex words containing prefixes/suffixes or irregular words e.g. *souvenir*, *destruction*, and *conscious*.	4
• Attempts to spell some complex or polysyllabic words using visual or phonetic strategies e.g. *'safariye'* for *safari*, '*adventerous*' for *adventurous*.	3
• Spells the majority of high frequency common words correctly e.g. *inside*, *because*, *while*.	2
• Spells some common monosyllabic words correctly (e.g. *mum*, *cat*, *bird*). Uses phonic strategies to attempt to spell high frequency common words e.g. '*grat*' for *great*, '*fhun*' for *fun*.	1
• Below the level for 1 mark.	0
Punctuation	Mark
• Uses a range of punctuation to clarify structure and create effect (e.g. speech marks, dashes, brackets, apostrophes, commas to demarcate sentences).	4
• Secure use of full stops and capital letters. Uses punctuation in addition to capital letters and full stops, the majority are used correctly (e.g. question marks, exclamation marks, commas in lists).	3
• Evidence of accurate use of capital letters and full stops, however few there are (e.g. sentence finishes with a full stop and next sentence begins with a capital letter).	2
• Shows awareness of how full stops are used in writing.	1
• Below the level for 1 mark.	0
Sentence Structure and Grammar	Mark
• Secure control of complex sentences. Understands how clauses can be manipulated for effect. Able to use conditional and passive voice.	4
• Beginning to write extended sentences including subordinators (e.g. *if, so, while, when, after*). The basic grammatical structure of sentences usually correct (e.g. usually consistent and correct use of tenses and nouns and verbs agree).	3
• Beginning to use other conjunctions to create compound sentences (e.g. *but*) and may be using multiple clauses (still mixing up tenses).	2
• Writes simple sentences which include the conjunction '*and*'.	1
• Below the level for 1 mark.	0
Vocabulary	Mark
• Demonstrates use of well-chosen, vivid and powerful vocabulary to create effect (e.g. verbs, adjectives, adverbs).	4
• Varied use of adjectives, verbs and specific nouns (e.g. *delicious* for *nice*/*sauntered* for *went/poodle* for *dog*).	3
• Some selection of interesting and varied verbs e.g. *jumped, compare, guess*.	2
• Uses simple vocabulary, appropriate to content. Writing is composed of simple nouns and verbs, e.g. *look, went, go, play, see*.	1
• Below the level for 1 mark.	0
Organisation and Overall Structure	Mark
• Paragraphs are well organised, based on themes and provides a cohesive text for the reader (e.g. paragraphs, subheadings, logically organised events).	4
• Uses paragraphs to organise writing, showing an identifiable structure. May be short sections.	3
• Themes are expanded upon and linked together in a series of sentences.	2
• Communicates meaning but may 'flit' from idea to idea and any themes that are expanded are done so in one sentence.	1
• Below the level for 1 mark.	0
Ideas	**Mark**
• Ideas are creative and interesting in a way that engages the reader. Uses a range of strategies and techniques such as asides, comment, observation, anticipation, suspense, tension.	4
• Ideas are imaginative and varied evidence of descriptive detail about characters, settings, feelings, emotions and actions.	3
• Ideas are developed to by adding detail (e.g. is beginning to provide additional information or description beyond a simple list).	2
• Produces short sections of ideas which may be repetitive and limited in nature.	1
• Below the level for 1 mark.	0

Appendix O: Baseline Proforma

HELPING HANDWRITING SHINE: Proforma for all schools - Spring 2018

The National Foundation for Educational Research (NFER) has been commissioned by the Education Endowment Foundation (EEF) to evaluate the impact of

Question 1: Does/will your school provide any **additional** literacy support for Year 2? *If the answer is 'yes', please complete questions 2-5. If the answer is 'no', please go to question 6.*	YES		NO	

Question 2: If yes;	NAME OF PROGRAMME	
	START DATE	
	END DATE	
	Any other details	

Question 3: If yes, is it for all pupils or targeted pupils who require additional support?	TARGETED		ALL	

Question 4: If targeted, how are these pupils identified; what is the criteria to receive this additional support?	
Question 5: Who (which staff member/s) provide the literacy support? E.g. Tas, teachers; do not name individuals	

Question 6: Does/will your school provide any **additional** literacy support for Year 5? *If the answer is 'yes', please complete questions 7-10. If the answer is 'no', you have finished the form.*	YES		NO	

Question 7: If yes;	NAME OF PROGRAMME	
	START DATE	
	END DATE	
	Any other details	

Question 8: If yes, is it for all pupils or targeted pupils who require additional support?	TARGETED		ALL	

Question 9: If targeted, how are these pupils identified; what is the criteria to receive this additional support?	
Question 10: Who (which staff member/s) provide the literacy support? E.g. Tas, teachers; do not name individuals	

Thank you for your support with this information. Please now upload this form to your NFER school portal, using your instructions in your school letter.

Appendix P: Logs

Log 1: Delivery log Year 2

	School name:			During the programme, you should update this log to show what Helping Handwriting Shine activities you have completed **for each class**. This can be done daily, or at the end of each week. It is important to be honest in your reflections. If you have any questions, contact handwriting@nfer.ac.uk								
	School DfE number:											
	Class name/number/ID:											
	Session no.	Date (DD/MM)	Session Delivery (Length of session)	Prepare - were pupils given 5 minutes to follow the 4 Ps? (Yes or No)	Warm-up activity delivered (1 to 6)	Writing activity delivered (1 to 5)	Tailored levels used? (Yes or No)	Model-plan-evaluate delivered? (Yes or No)	Extension activity delivered? (Yes or No)	How many gap tasks did you deliver after the session? (0-2+)	Please rate class engagement in this session	Please describe any adaptations made
Week 1	Session 1											
	Session 2											
	Session 3											
Week 2	Session 4											
	Session 5											
	Session 6											
Week 3	Session 7											
	Session 8											
	Session 9											
Week 4	Session 10											
	Session 11											
	Session 12											
Week 5	Session 13											
	Session 14											
	Session 15											

Log 2: Reflections log Year 2

		Select your answer	Comments
1	On average, how frequently did you make **adaptations** to the programme? (for e.g. changing the content, order or activities; purposefully missing anything)		
2	Please indicate any approximate **costs** you have incurred in delivering the programme (for e.g. travel costs, photocopying, materials)		
3	On average, how much **time,** outside of teaching, did you spend per week on the intervention? (for e.g. reading, planning or marking, other than what would normally be done)		
4	Describe any particular **issues or challenges** you faced during delivery (for e.g. time, resources, space; anything to do with the content of the programme)	N/A	
5	**Any other reflections**	N/A	

Log 3: Post programme log Year 2

	To what extent have you / your school...	Have you/your school...			
	...continued to use the activities, techniques and/or gap tasks from programme?	...requested/received any support from the Leeds team?	...received any other external support or CPD for handwriting?	...received any other external support or CPD for literacy?	Comments
February					
March					
April					
May					
June					

Log 1: Delivery log Year 5

	School name:		RPO INSERT SCHOOL NAME HERE		During the programme, you should update this log to show what Helping Handwriting Shine activities you have completed **for each pupil**. Use the drop down box in cell D/E5 to select the pupil. This can be done daily, or at the end of each week. It is important to be honest in your reflections. If you have any questions, contact handwriting@nfer.ac.uk						
	School DfE number:		RPO INSERT DFE NO HERE								
	Pupil name:		INSERT DROP DOWN HERE								
	Session no.	1. Date (DD/MM)	2. Session Delivery (Length of session)	3. Prepare - was the pupil given 5 minutes to follow the 4 Ps? (Yes or No)	4. Warm up		5. Writing		6. Model-plan-evaluate delivered? (Yes or No)	7. Did the pupil complete the extension activity? (Yes or No)	8. Please describe any adaptations made
					a) Which activity (1 to 5)	b) At what level (1 to 3)	a) Which activity (1 to 6)	b) At what level (1 to 3)			
Week 1	Session 1										
	Session 2										
	Session 3										
Week 2	Session 4										
	Session 5										
	Session 6										
Week 3	Session 7										
	Session 8										
	Session 9										
Week 4	Session 10										
	Session 11										
	Session 12										
Week 5	Session 13										
	Session 14										
	Session 15										
Week 6	Session 16										
	Session 17										
	Session 18										

Log 2: Reflections log Year 5

	Helping Handwriting Shine: Y5 Overall Reflections	Please complete this log at the end of the intervention period i.e. after the eight weeks of programme delivery has been completed. Click on the drop down boxes and select your answer. Leave any comments or description in the next column.	
	School name:		
	School DfE Number:		
	Class name/number/ID:		
		Select your answer	Comments
1	On average, how frequently did you make **adaptations** to the programme? (for e.g. changing the content, order or activites; purposefully missing anything)		
2	Please indicate any approximate **costs** you have incurred in delivering the programme (for e.g. travel costs, photocopying, materials)		
3	On average, how much **time,** outside of teaching, did you spend per week on the intervention? (for e.g. reading, planning or marking, other than what would normally be done)		
4	**Contamination:** Were there any occasions where Year 5 pupils not receiving the programme, were accidentally (or otherwise) exposed to the programme? (for e.g. did they hear about or see any Helping Handwriting Shine work or teaching?)		
5	Describe any particular **issues** or **challenges** you faced during delivery (for e.g. time, resources, space; anything to do with the content of the programme)	N/A	
	Any other reflections	N/A	

Log 3: Post-programme log Year 5

Helping Handwriting Shine RCT: Year 5 Monthly Post-Programme Log						
School name:		Once you have finished delivery of the eight-week programme, please use this log to give a monthly overview of you/your school's activities for Year 5, until testing in June. Select from the drop down boxes then give additional details in the 'comments' column.				
School DfE number:						
Class name/number/ID:						
	To what extent have you / your school...	**Have you/your school...**			Are you aware of any contamination between the control and intervention groups? (for e.g. control group pupils hear about or see any Helping Handwriting Shine work or teaching?)	Comments
	...continued to use the activities, techniques and/or gap tasks from programme, with any pupil?	...requested/received any support from the Leeds team?	...received any other external support or CPD for handwriting?	...received any other external support or CPD for literacy?		
February						
March						
April						
May						
June						

Business As Usual Log – Year 2

	Helping Handwriting Shine: Control School Log - Year 2					
	Year 2 teachers should complete this log, reflecting on teaching and learning for their Year 2 class during the period 1 November 2018 to 31 January 2019. Once complete, upload the log to the online portal. Email handwritingRCT@nfer.ac.uk for any queries					
	School name:				**Responses** Select a response from the drop down box	**Comments**
	School DfE number:					
	Class name/number/ID:					
	Reflect on the period from 1 November 2018 to 31 January 2019. During this period:					
1	Did you have allocated time specifically for handwriting practice in class? If so, how often?					
2	If you **did** have specific handwriting practice sessions:	a) How long were they, on average?				
		b) What activities did they usually entail? (Select up to 3)	1			
			2			
			3			
		c) Were tasks and activities usually individualised?				
3	Did you use any specific **handwriting** interventions or programmes during this period?					
4	Did you use any specific **literacy** interventions or programmes during this period?					
5	Did any pupils receive literacy or handwriting support, in addition to normal teaching?					

Business As Usual log – Year 5

Helping Handwriting Shine: Control Pupil Log - Year 5

Year 5 teachers should complete this log to reflect on the teaching and learning for pupils who were selected as eligible to receive the intervention, but were allocated to the control group. These are the pupils who have NOT received the intervention, and were taught by the normal class teacher. Please reflect on teaching and learning for these pupils during the period 1 November 2018 to 31 January 2019.

				Responses Select a response from the drop down box	Comments
	School name:				
	School NFER number:				
	Reflect on the period from 1 November 2018 to 31 January 2019. During this period:				
1	Was there time specifically allocated to handwriting practice for these pupils? If so, how often?				
2	If you **did** have specific handwriting practice sessions:	a) How long were they, on average?			
		b) What activities did they usually entail? (Select up to 3)	1		
			2		
			3		
		c) Were tasks and activities usually individualised?			
3	Did you use any specific **handwriting** interventions or programmes for these pupils, during this period?				
4	Did you use any specific **literacy** interventions or programmes for these pupils, during this period?				
5	Did any of these pupils receive literacy or handwriting support, in addition to normal teaching?				

Appendix Q1: Writing test for Year 2 Experiment

School Trip

Imagine that you could go anywhere you wanted to on a school trip with your class and your teacher.

You could go anywhere at all.

Write about where you would go and what you would do.

Appendix Q2: Writing test for Year 5 Experiment

School Trip

Imagine that you could go anywhere you wanted to on a school trip with your class and your teacher.

You could go anywhere at all.

Write about where you would go and what you would do.

Appendix Q3: Administration instructions for writing test (both experiments)

Administration Instructions for writing test

Y2 / Y5 writing (20 minute task)

- Resources: writing booklet, agreed writing implement
- Instructions to teacher / administrator:
  - Ensure each pupil has the writing booklet and pencil or pen (whatever has been agreed)
  - Explain to the class / group that they are going to do some writing for **20 minutes** and that you will tell them what they will be writing about, when to start writing and when to stop writing.
  - Read the task instruction out loud to the class, whilst they follow in their booklets where the words are written – ie ***Imagine that you could go anywhere you wanted to on a school trip with your class and your teacher. You could go anywhere at all. Write about where you would go and what you would do.***
  -
  - Tell the children to start their writing.
  - After 15 minutes, tell the children that the time is nearly finished, so they should carry on with the writing but remember that they only have about 5 minutes left.
  - **After 20 minutes**, tell the children to **stop** writing.

Appendix R1: Handwriting test for Year 2 Experiment

Handwriting task

The quick brown fox jumps over the lazy dog.

Appendix R2: Handwriting test for Year 5 Experiment

Handwriting

Copy out the sentence in the box as many times as possible. Your teacher will tell you when to start writing and when to stop writing.

The quick brown fox jumps over the lazy dog.

Appendix R3: Administration instructions for Handwriting test (both experiments)

Administration Instructions for handwriting test

Y2 / Y5 handwriting (3 minute task)

- Resources: handwriting page of booklet, agreed writing implement, stopwatch for administrator

- Instructions to teacher / administrator:
  - Explain to the pupils that they will be given a test to see how quickly and neatly they can write and that it is important to listen carefully to all the instructions given to them.
  - Distribute handwriting page of booklet and ensure they have agreed writing implement.

  - Read out these instructions:
    "Please do not pick up your pencil or write anything until I say so.
    When I say 'go', I'd like you to write the sentence: The quick brown fox jumps over the lazy dog. The sentence is at the top of the page.
    Keep writing it over and over until I say 'stop' – you'll be writing it for 3 minutes.
    Do not rub out or cross out any mistakes. It doesn't matter if you start a new line for each sentence, or not.
    Remember it is not a race, just use your normal writing.
    Write as quickly but as neatly as you can until I say stop"

  - When the pupils appear ready, say: "***ready – go***".

  - Start stopwatch and watch carefully that the pupils keep repeating the sentence.

- After one minute, say: "***Two minutes to go – remember to write as quickly but as neatly as you can"***
- After two minutes, say: "***One minute to go – keep writing"***
- After three minutes: "***Stop writing, please put your pencils down now"***

Ensure writing has stopped. Watch carefully that no further writing occurs. Collect booklets.

Appendix S1: Randomisation and Sampling Codes

Year 2 Experiment Randomisation

```
## EEFH School Randomisation-Stratified by Training Hub

#1. Set work directory

setwd("…")


#2.identify project

project<-"EEFH"


#3.identify classification: c, r or p

classification<-"r"


#4. Number of the randomisation: 1st, 2nd, 3rd ...

randomisation<-1

randomisation<-as.character(as.roman(randomisation))


###5. Load data

Experiment<-read.csv("…",header=T,stringsAsFactors = F)


###Identify stratification and unique identifier variables

#6.list the stratification variables

stratification<-list("Hub")

n_strats<-length(stratification)


#7.unique school identifier variable

ui<-"NFER_No"


###8. What time is now? (hh.mm)

time_now<-21.40
```

```
aux<-100*trunc(time_now)+100*(time_now-trunc(time_now))

set.seed(aux)

seeds<-sample(1:9999,size=(n_strats+2))


#Duplicated ui information and lines with no ui identification

#must removed

Experiment<-Experiment[!duplicated(Experiment[ui]),]

Experiment<-Experiment[!is.na(Experiment[ui]),]


#Keep the original order of the columns

originalColOrder<-colnames(Experiment)


###Adding a variable that will allow for the recovery

##of the original oder of the data frame rows later on

Experiment$originalRowOrd<-1:nrow(Experiment)


### Ordering Experiment by unique identifier

Experiment<-Experiment[order(Experiment[ui]),]


### Assigning a random order to the stratification

rands<-paste("rand",as.character(1:n_strats),sep="_")


for (i in 1:n_strats){


aux<-as.data.frame(sort(unique(Experiment[,stratification[[i]]])))

set.seed(seeds[1])

seeds<-seeds[-1]


aux[rands[i]]<-sample(1:nrow(aux))


Experiment<-merge(Experiment,aux,by.x=stratification[[i]],by.y=colnames(aux)[1])
```

```
}

###Randomise by unique identifier
set.seed(seeds[1])
seeds<-seeds[-1]
Experiment["rand_ui"]<-sample(nrow(Experiment))

###Reorder the rows of Experiment by rands and rancluster
rands<-c(rands,"rand_ui")
aux<-do.call(order,Experiment[rands])
Experiment<-Experiment[aux,]

###Assigning Control or Intervention Group
aux<-rep(1:2,times=round(nrow(Experiment)/2))
Experiment$grp<-aux[1:nrow(Experiment)]

rands<-c(rands,"grp")

aux<-data.frame(group=c("control","intervention"))
set.seed(seeds[1])
aux$randgroup<-sample(1:2)

Experiment<-merge(Experiment,aux,by.x="grp",by.y="randgroup")

##Returning the data frame to its original order
Experiment<-Experiment[order(Experiment$originalRowOrd),]

###Removing the variables that are no longer necessary
originalColOrder<-c(originalColOrder,"group")
Experiment<-Experiment[,originalColOrder]
```

Year 2 Experiment Secondary Sample

```
### EEFH -Sample a pre-defined number of cases from each

## cluster


#1. Set work directory

setwd("…")


#2.identify project

project<-"EEFH"


#3.identify classification: C, R or P

classification<-"R"


###4. Load data

Sample<-read.csv("…",header=T,stringsAsFactors = F)


#5.identify the clustering variable: School

cluster<-colnames(Sample)[1]


###Eliminate the cases with no assigned school

aux<-nrow(Sample)

Sample<-Sample[complete.cases(Sample[cluster]),]


###No cases were lost

aux-nrow(Sample)==0


###Count the number of schools:103

clusts<-unique(Sample[cluster])

n_clusters<-nrow(clusts)


##6.Define the variable from where the cases will be drawn from
###In this case, pupils in schools
```

```
cases<-colnames(Sample)[2]

###Eliminate the cases for which the cases variables is not assigned

Sample<-Sample[complete.cases(Sample[cases]),]

###7. Define the number of cases to pick from each cluster
pick<-5

###8. What time is now? (hh.mm)
time_now<-18.11

####Safety check? Do all schools have at least 5 elements: yes, all good!
aux<-as.data.frame(table(Sample$NFER_No))
min(aux$Freq)>4

aux<-100*trunc(time_now)+100*(time_now-trunc(time_now))
set.seed(aux)
seeds<-sample(1:9999,size=(n_clusters))



func<-function(i){
 char<-clusts[i,1]
 aux<-Sample[Sample[cluster]==char,cases]
 aux<-sort(unique(aux))
 set.seed(seeds[i])
 aux<-sample(aux,pick)
 aux<-sort(aux)
 return(aux)
}

putout<-lapply(1:n_clusters,func)
```

```
putout<-setNames(putout,clusts[,1])

putout<-do.call(rbind,putout)

putout<-putout[order(rownames(putout)),]
```

Year 5 Experiment Randomisation

```
## Code for EEFH's Year 5 pupil level randomisation

#1. Set work directory

setwd("…")


#2.identify project

project<-"EEFH"


#3.identify classification: c, r or p

classification<-"r"


#4. Number of the randomisation: 1st, 2nd, 3rd ...

randomisation<-2

randomisation<-as.character(as.roman(randomisation))


###5. Load data

Experiment<-read.csv("…",header=T,stringsAsFactors = F)


###Identify stratification and randomisation variables


#6.list the stratification variables

aux<-colnames(Experiment)

### aux[1] corresponds to the school identifier and aux[9] to pupils' reported FSM status
stratification<-list(aux[1],aux[9])

n_strats<-length(stratification)


#7.identify the randomisation/cluster variable

##aux[3] corresponds to the pupil identifier
aucluster<-aux[3]


###8. What time is now? (hh.mm)

time_now<-13.07
```

```
aux<-100*trunc(time_now)+100*(time_now-trunc(time_now))

set.seed(aux)

seeds<-sample(1:9999,size=(n_strats+2))


#Duplicated cluster information and lines with no cluster identification

#must removed

#No lines were removed

Experiment<-Experiment[!duplicated(Experiment[cluster]),]

Experiment<-Experiment[!is.na(Experiment[cluster]),]


#Keep the original order of the columns

originalColOrder<-colnames(Experiment)


###Adding a variable that will allow for the recovery

##of the original oder of the data frame rows later on

Experiment$originalRowOrd<-1:nrow(Experiment)


### Ordering Experiment by cluster

Experiment<-Experiment[order(Experiment[cluster]),]


### Assigning a random order to the stratification

rands<-paste("rand",as.character(1:n_strats),sep="_")


for (i in 1:n_strats){


aux<-as.data.frame(sort(unique(Experiment[,stratification[[i]]])))

set.seed(seeds[1])

seeds<-seeds[-1]


aux[rands[i]]<-sample(1:nrow(aux))
```

```
Experiment<-merge(Experiment,aux,by.x=stratification[[i]],by.y=colnames(aux)[1])
}

###Randomise by cluster/randomisation variable
set.seed(seeds[1])
seeds<-seeds[-1]
Experiment["rand_cluster"]<-sample(nrow(Experiment))

###Reorder the rows of Experiment by rands and rancluster
rands<-c(rands,"rand_cluster")
aux<-do.call(order,Experiment[rands])
Experiment<-Experiment[aux,]

###Assigning Control or Intervention Group
aux<-rep(1:2,times=round(nrow(Experiment)/2))
Experiment$grp<-aux[1:nrow(Experiment)]

rands<-c(rands,"grp")

aux<-data.frame(group=c("control","intervention"))
set.seed(seeds[1])
aux$randgroup<-sample(1:2)

Experiment<-merge(Experiment,aux,by.x="grp",by.y="randgroup")

##Returning the data frame to its original order
Experiment<-Experiment[order(Experiment$originalRowOrd),]

###Removing the variables that are no longer necessary
rands<-c("originalRowOrd",rands)
rands<-which(colnames(Experiment)%in%rands)
Experiment<-Experiment[,-rands]
```

```
originalColOrder<-c(originalColOrder,"group")

Experiment<-Experiment[,originalColOrder]
```

Appendix S2: Analysis Code (Extracts)

Year 2 Experiment Primary Analysis

```
###Primary ITT Analysis for EEFH Y2

(The data for the analysis has previously been cleaned and uploaded)


###Primary outcome

aux<-which(colnames(Y2)=="E_WAMCJ")

###Randomisation group

aux<-c(aux,which(colnames(Y2)=="Rand_S"))

##Baseline

aux<-c(aux,which(colnames(Y2)=="B_WAMCJ"))

## Stratifier

aux<-c(aux,which(colnames(Y2)=="Region"))

###Cluster

aux<-c(aux,which(colnames(Y2)=="School"))


####Selecting cases to go into the model

aux1<-Y2$IN_Analysis=="yes"

Df<-Y2[aux1,aux]

colnames(Df)<-c("outcome","treatment","baseline","strat","cluster")


MOD<-lmer(outcome~treatment+baseline+strat+(1|cluster),REML=FALSE, data=Df)


###Testing the model

addWorksheet(wb, "EEFH_Y2_Primary")

simulationOutput <- simulateResiduals(fittedModel = MOD)

plot(simulationOutput)

testUniformity(simulationOutput = simulationOutput)
```

Year 5 Experiment Primary Analysis

###Primary ITT Analysis for EEFH Y5

(The data for the analysis has previously been cleaned and uploaded)

###Primary outcome

aux<-which(colnames(Y5)=="E_WAMCJ")

###Randomisation group

aux<-c(aux,which(colnames(Y5)=="Rand_P"))

##Baseline

aux<-c(aux,which(colnames(Y5)=="B_WAMCJ"))

Stratifier1

aux<-c(aux,which(colnames(Y5)=="Rep_FSM"))

Stratifier2

aux<-c(aux,which(colnames(Y5)=="School"))

Df<-Y5[,aux]

colnames(Df)<-c("outcome","treatment","baseline","strat1","strat2")

###Retaining just complete cases

aux<-complete.cases(Df)

###Removing empty levels

Df$treatment<-droplevels(Df$treatment)

Df$strat1<-droplevels(Df$strat1)

Df<-Df[aux,]

###This model follows the approach OLS+stratifiers as dummies

MOD1<-lm(outcome~treatment+baseline+strat1+strat2,data=Df)

```
aux<-summary(MOD1)$coefficients
```

Year 2 Experiment Secondary Analysis A

###Secondary ITT A Analysis for EEFH Y2

(The data for the analysis has previously been cleaned and uploaded)

```
###Secondary Outcome A
aux<-which(colnames(Y2)=="E_WAMCR")
###Randomisation group
aux<-c(aux,which(colnames(Y2)=="Rand_S"))
##Baseline
aux<-c(aux,which(colnames(Y2)=="B_WAMCJ"))
## Stratifier
aux<-c(aux,which(colnames(Y2)=="Region"))
###Cluster
aux<-c(aux,which(colnames(Y2)=="School"))

Df<-Y2[,aux]
colnames(Df)<-c("outcome","treatment","baseline","strat","cluster")

###Retaining just complete cases
aux<-complete.cases(Df)

Df<-Df[aux,]

MOD<-lmer(outcome~treatment+baseline+strat+(1|cluster),REML=FALSE, data=Df)

###Testing the model
addWorksheet(wb, "EEFH_Y2_SecondaryA")
simulationOutput <- simulateResiduals(fittedModel = MOD)
plot(simulationOutput)
testUniformity(simulationOutput = simulationOutput)
```

```
insertPlot(wb,"EEFH_Y2_SecondaryA")
```

Year 5 Experiment Secondary Analysis A

```
###Secondary ITT  A Analysis for EEFH Y5

(The data for the analysis has previously been cleaned and uploaded)

###Secondary Outcome A

aux<-which(colnames(Y5)=="E_WAMCR")

###Randomisation group

aux<-c(aux,which(colnames(Y5)=="Rand_P"))

##Baseline

aux<-c(aux,which(colnames(Y5)=="B_WAMCJ"))

## Stratifier1

aux<-c(aux,which(colnames(Y5)=="Rep_FSM"))

## Stratifier2

aux<-c(aux,which(colnames(Y5)=="School"))



Df<-Y5[,aux]

colnames(Df)<-c("outcome","treatment","baseline","strat1","strat2")

###Retaining just complete cases

aux<-complete.cases(Df)

Df<-Df[aux,]

MOD1<-lm(outcome~treatment+baseline+strat1+strat2.,data=Df)
```

Year 2 Experiment Secondary Analysis B

```
###Secondary ITT  B Analysis for EEFH Y2

(The data for the analysis has previously been cleaned and uploaded)


###Secondary Outcome B

aux<-which(colnames(Y2)=="E_HSTRSfac")

###Randomisation group

aux<-c(aux,which(colnames(Y2)=="Rand_S"))

##Baseline

aux<-c(aux,which(colnames(Y2)=="B_HSTRSfac"))

## Stratifier

aux<-c(aux,which(colnames(Y2)=="Region"))

###Cluster

aux<-c(aux,which(colnames(Y2)=="School"))


Df<-Y2[,aux]

colnames(Df)<-c("outcome","treatment","baseline","strat","cluster")


### Checking the structure of the data. I need to convert the ouctome and baseline_fac variables into factors

str(Df)

outcome<-sort(unique(Df$outcome))

Df$outcome<-factor(Df$outcome, levels=outcome, ordered=TRUE)

baseline<-sort(unique(Df$baseline))

Df$baseline<-factor(Df$baseline, levels=baseline, ordered=TRUE)


Df<-Df[complete.cases(Df),]

library(ordinal)

(Two models were run, the logit was chosen for the analysis  because it had the lower Akaike information criteria (AIC) and lower absolute values of log-likelihood (logLik))

logit_model<-clmm(outcome~treatment+baseline+strat+(1|cluster),data=Df,Hess=T,nAGQ=15)

probit_model<-clmm(outcome~treatment+baseline+strat+(1|cluster),data=Df,link = "probit",Hess=T,nAGQ=15)

summary(logit_model)
summary(probit_model)
```

MOD<-logit_model

Year 5 Experiment Secondary Analysis B

###Secondary ITT B Analysis for EEFH Y5

(The data for the analysis has previously been cleaned and uploaded)

###Secondary Outcome B

aux<-which(colnames(Y5)=="E_HSTRSfac")

###Randomisation group

aux<-c(aux,which(colnames(Y5)=="Rand_P"))

##Baseline

aux<-c(aux,which(colnames(Y5)=="B_HSTRSfac"))

###Stratifier 1

aux<-c(aux,which(colnames(Y5)=="Rep_FSM"))

Stratifier 2

aux<-c(aux,which(colnames(Y5)=="School"))

Df<-Y5[,aux]

colnames(Df)<-c("outcome","treatment","baseline","strat1","strat2")

Checking the structure of the data. I need to convert the ouctome and baseline_fac variables into factors

str(Df)

outcome<-sort(unique(Df$outcome))

Df$outcome<-factor(Df$outcome, levels=outcome, ordered=TRUE)

baseline<-sort(unique(Df$baseline))

Df$baseline<-factor(Df$baseline, levels=baseline, ordered=TRUE)

Df<-Df[complete.cases(Df),]

library(ordinal)

(Several models were run, the cloglog was chosen for the analysis because it had the lower Akaike information criteria (AIC) and lower absolute values of log-likelihood (logLik))

```
logit_model<-clm(outcome~treatment+baseline+strat1+strat2,data=Df)

probit_model<-clm(outcome~treatment+baseline+strat1+strat2,data=Df,link = "probit")

loglog_model<-clm(outcome~treatment+baseline+strat1+strat2,data=Df,link = "loglog")

cloglog_model<-clm(outcome~treatment+baseline+strat1+strat2,data=Df,link = "cloglog")

cauchit_model<-clm(outcome~treatment+baseline+strat1+strat2,data=Df,link = "cauchit")


summary(logit_model)

summary(probit_model)

summary(loglog_model)

summary(cloglog_model)

summary(cauchit_model)

MOD<-cloglog_model
```

This document is available for download at **https://educationendowmentfoundation.org.uk**

Education Endowment Foundation

The Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP

https://educationendowmentfoundation.org.uk

@EducEndowFoundn

Facebook.com/EducEndowFoundn