FLASH Marking

Evaluation Report

April 2022

Independent Evaluators:

Dr Rebecca Morris, Professor Stephen Gorard, Professor Beng Huat See, Dr Nadia Siddiqui.

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

FLASH Marking was evaluated by a team from Durham University and the University of Warwick: Rebecca Morris, Stephen Gorard, Beng Huat See, and Nadia Siddiqui.

Contact details for correspondence:

Dr Rebecca Morris
Department of Education Studies, University of Warwick, Coventry, CV4 7AL
Email: rebecca.e.morris@warwick.ac.uk

Executive summary

The project

FLASH (Fast Logical Aspirational Student Help) Marking is a school-developed approach, from Meols Cop Research School in Southport, in which teachers give skills-based feedback rather than grades in Key Stage 4 English. Skills required to access the top band of GCSE English and English Literature performance are translated into short codes that teachers use when marking work. These codes are also used in lessons to teach the skills and for teachers to track strengths and weaknesses across their classes and by pupils to make peer- and self-marking faster, focused, and more useful. The intervention was designed to improve attainment in Key Stage 4 English Language and Literature and to reduce workload (particularly from marking) for teachers.

A total of 18,500 Key Stage 4 pupils (Years 10 and 11; aged 15–16) from 103 schools took part in this two-year trial (2018–2020), which was delivered by trained teachers and supported by teaching assistants (TAs) and senior leaders. Pupils were also taught to use the coding in their self-assessment and in peer marking as well as through other teaching and learning activities: for example, FLASH codes often being shared as learning objectives or outcomes on resources such as PowerPoints and worksheets.

This trial was designed as an efficacy trial. However, this project and its evaluation were affected by 2020 partial school closures caused by the Covid-19 pandemic, and the cancellation of GCSE examinations in 2020. As a result, the evaluators were not able to use GCSEs to estimate the impact of the project on English Language and Literature attainment and therefore, it was not possible to rate the security of impact estimates. This report focuses on the secondary outcome of teacher workload, reported through pre- and post-intervention surveys, and an in-depth implementation and process evaluation involving 16 case study schools, observations of lessons and interviews with school staff (prior to Covid-19), observations of training sessions, and information from teacher and pupil questionnaires.

The delivery of FLASH Marking intervention to pupils started in September 2018 (teacher training took place in June–July 2018). Delivery ran through until summer 2020, although in many schools teaching and learning was severely disrupted by Covid-19 between March–July 2020 and FLASH Marking use was reduced or discontinued in some schools.

FLASH Marking was developed by two senior leaders originally based at Meols Cop Research School and is now run as an independent company: https://flashmarking.com/.

Table 1: Key conclusions

Key Conclusions
1. Due to Covid-19 the primary outcome for this evaluation (GCSEs) was not collected and so no measure of impact on English Language and Literature attainment is reported and there is no security rating for this trial. Key conclusions are based on qualitative data from the implementation and process evaluation.
2. Teachers receiving FLASH Marking reported a greater reduction in both total hours spent working and hours spent marking compared to teachers who did not receive the intervention. This was reported by teachers who completed both workload surveys—pre- and post-intervention—due to dropout, this was approximately 25% of the number of teachers who responded to the initial survey.
3. Process evaluation findings indicate that attendance at training was high, and the intervention was generally received positively by teachers. FLASH Marking was implemented as intended in most schools with some variation, mostly due to staff engagement. In total, 90% of English department Heads responding (n=40) to the second workload survey strongly agreed or tended to agree that their department had been fully committed to FLASH Marking since the start of the trial.
4. As observed in training sessions, case studies, and staff questionnaires, where leadership personnel changed, a school's emphasis on and engagement with FLASH Marking sometimes changed too. Where senior leaders were also members of the English department, there tended to be increased buy-in to FLASH Marking.
5. Teachers were pleased with the quality of training and the support available from the delivery team. Discussions of good practice across school hub groups were received enthusiastically, particularly in the earlier stages of the trial. 14 schools reported that they had cascaded the FLASH Marking training to other subject departments in their schools.

Additional findings

For reported teacher workload (secondary outcome), schools receiving the intervention reported a greater reduction in time spent marking than control group schools (effect size -0.17). This was based on 218 matched responses to pre- and post-intervention surveys: the first survey received 833 teacher responses at the outset of the trial. The outbreak of the Covid-19 pandemic meant that the post-intervention response was far lower than would normally be expected and therefore, these findings are based on a smaller sample.

In the post-intervention survey, 159 teachers from the intervention group largely tended to agree that the training they received to use FLASH Marking was helpful, it had reduced their marking workload, and that they would encourage other schools to use it. In total, 77% of these teachers reported that the intervention had benefited pupils.

Teachers (and some pupils) reported mixed views about the extent to which FLASH Marking was suitable for groups with different levels of prior attainment. Some teachers felt that FLASH worked well with all learners, while others were concerned that it was not always appropriate for those with lower prior attainment. A small number of higher attaining learners in case study schools reported wanting more in-depth written feedback and tended to find FLASH too neutral in its approach.

A number of English departments reported using FLASH Marking with Key Stage 3 groups (in addition to the Key Stage 4 trial cohort). By September 2019, 31 of 52 schools had rolled-out the programme to all of their Key Stage 3 year groups and a total of 14 schools reported that other subject departments (such as History, Modern Languages, Science, and Media Studies) had introduced FLASH Marking too. Teachers were very positive about the potential for FLASH to positively impact attainment and staff workload with these year groups, indicating that there are possibilities for use with different age groups and for subject-specific adaptation of the intervention.

Some aspects of the logic model were supported by data from the implementation and process evaluation (IPE). The training sessions (three whole-day sessions held at different timepoints throughout the trial, for two colleagues from each school) had high attendance. These and additional support from the development team were found to have a positive impact on engagement and delivery of FLASH Marking.

Through visits to case study schools and incorporating information from teacher and pupil surveys, and correspondence with the development team, four active ingredients were identified as playing a central role in the implementation of FLASH Marking. These elements were shared explicitly with English department staff at each of the training sessions and reiterated by the development team as part of their follow-up support work.

- Removal of number/letter-based grades from day-to-day feedback.
- Using FLASH codes as part of English lessons, for example in success criteria, learning outcomes, and model examples.
- FLASH codes are used by teachers to provide personalised feedback to pupils and to inform planning of lessons.
- FLASH codes are used to support pupils' metacognition, for example through reflection and improvement tasks, and target-setting for future work.

During the second year of the trial, some English departments did appear to be using number/letter-based grades more frequently than previously, as highlighted by pupil survey responses. Interviews with teachers indicated that while they were generally pleased to have shifted the focus to more formative feedback approaches, most acknowledged that removing grades altogether was sometimes not easy or feasible in their current contexts.

Findings related to the workload outcome build on previous studies in this area and indicate the potential for workload reduction as a result of changes to marking policies and practices. Development of a more reliable measure of teacher workload would be beneficial for future studies.

Cost

The average cost of FLASH Marking for one school was around £7.96 per pupil, per year, when averaged over three years. Over the course of three years, this equates to £23.89 per pupil. Schools are required to release two English teachers (including the Head of Department) to attend three training days. Schools are also expected to accommodate a support and monitoring visit from the FLASH Marking development team each year.

Introduction

Background

Over the past 30 years, a growing body of research has pointed to the value of providing pupils with high-quality formative feedback to improve progress and attainment (Black and William, 1998; Hattie and Timperley, 2007; Newman *et al.*, 2021). When done well, feedback can help pupils to understand their strengths and areas for development, and can help them to address where improvements are needed in order to move their learning forward. Despite recognising the importance of feedback for learning, there is still an overall lack of strong research evidence on the most effective approaches and methods that school leaders and teachers might choose to employ. Recent reviews have particularly highlighted the lack of large-scale, high-quality studies that examine the use of written feedback on pupils' work (Elliott *et al.*, 2016; Newman *et al.*, 2021). Elliott *et al.* (2016) note that marking is just one form of feedback provision, but it is often viewed as the main way by which pupils receive written feedback on their work. It can include grades, numbers, comments, or other symbolic methods (for example, codes, images, and symbols). Here, we briefly summarise some of the available evidence on marking and written feedback, providing a context for understanding the development and evaluation of the FLASH (Fast Logical Aspirational Student Help) Marking intervention.

Recent research provides a mixed picture on the impact of grading on pupil attainment. Some studies have indicated that grade-only feedback can have a less positive impact on attainment than approaches, which include some element of process or strategy comments (Elliott *et al.*, 2016; Newman *et al.*, 2021). A longitudinal study in Sweden compared the outcomes of 8,558 children in Grades 7 to 9, half of whom received attainment grades in Grade 6 and half who did not (Klapp, 2015). This was a natural experiment where the introduction of a new curriculum in 1969 allowed schools to decide whether to provide marks to children in Grade 6. The study reported that boys and lower attaining pupils did less well when receiving both grades and comments whereas there was a positive long-term effect for girls, particularly higher attainers. A review by Elliott *et al.* (2016) found that the regular grading of pupils' work was a common practice, particularly in secondary schools where 62% of survey respondents (n=679) reported grading or putting a mark on 'all', 'most' or 'some' of pupils' work. In primary schools, this figure was much lower at just 24%, and 51% of primary teachers (n=703) stating that they never used this practice.

The use of a combination approach (of comments and grades) has also produced unclear results in relation to pupils' attainment. In a small quasi-experimental study, Zhang and Misiak (2015) divided three Science classes in Grades 7 and 8 into each of three grading conditions (point-based, rubric-based, and rubric plus written feedback). Their study found that a combination of both written comments and grades was most effective for promoting pupil attainment. The authors also observed that written feedback, detailing specific areas for pupils to develop and work on, enabled pupils to reach the desired standards more quickly than if they were just given point-based or rubric-based feedback. Another study, with a cohort of college-level students rather than pupils in schools, found similar results, highlighting the strong relationship between detailed feedback and improved student grades (Lipnevic and Smith, 2009). This study also found that the descriptive feedback was more effective when given alone, unaccompanied by grades or praise.

The studies above, highlight the complexity of researching topics relating to written 'marking' and 'feedback'. In relation to both policy and practice, both terms hold varied meanings for school leaders and teachers. As such, implementation of feedback provision to children and young people in schools is diverse and sometimes difficult to accurately 'capture' through research (Elliott *et al.*, 2016; Elliott *et al.*, 2020). In addition to the practices of providing marks/grades and comments about pupils work, there are also approaches, which involve: indicating pupils' mistakes (sometimes requiring them to correct their work); acknowledgement marking (for example, ticking to show that something has been seen); dialogic marking (requiring pupils to write a response to a teachers' comment); and target-based approaches (see Elliott *et al.*, 2016 for further details). For years, teachers have also used different types of code-based marking to signify areas of success/development for pupils. Common examples include where 'Sp' is used to indicate a 'spelling' mistake on a piece of work or using // to suggest that a new paragraph is needed. Elliott *et al.* (2020) found that this approach continues to be widely used in English schools across the country. As far as we are aware, however, there are no rigorous, school-based studies, which examine the use of these code-based approaches nor the impact of more sophisticated code-based strategies, which support strategy or process (for example, promoting pupils self-assessment of their work, or supporting them to set targets and improve) (see also Newman *et al.*, 2021).

While noted as a relatively cost-effective approach for promoting achievement, feedback and marking do still incur considerable costs in relation to time and teacher resources (EEF, 2021a). Recent concerns around teacher workload have often centred on the excessive marking and feedback, which some teachers are required to engage with (Elliott *et al.*, 2020; Independent Workload Review Group, 2016). Some studies have also linked teacher workload (and specifically associated issues of marking/feedback) with negative impacts on teacher satisfaction and retention (Perryman and Calvert, 2020; Toropova *et al.*, 2021). As a result, there has been a growing interest, supported by government, Ofsted (Office for Standards in Education), and teacher unions, in finding more efficient and effective methods for providing high-quality feedback, which can still promote pupil learning. This has led to some innovative and potentially promising school-led approaches (see for example, Churches, 2020; Kime, 2018) but more evidence is needed on the value of these before they can be rolled-out more widely.

The FLASH Marking intervention was developed by school leaders at the Meols Cop Research School in 2015–2016. After some success in the Research School setting and some initial piloting with local schools, to test the implementation of the approach, an efficacy trial was planned. This was designed to be one of the largest trials of feedback approaches in schools conducted in the UK, aiming to examine the impact of FLASH on pupil attainment at the end of Year 11 and on teacher workload. The Covid-19 pandemic, unfortunately, meant that it was not possible to conclude the trial and to collect externally assessed GCSE data. As a result, we were not able to conduct analysis on attainment or share findings on this primary outcome. Nevertheless, this report presents analysis on the workload element of the study and provides findings from the in-depth implementation and process evaluation (IPE).

Intervention

Introduction: FLASH Marking is a school-developed feedback approach in which teachers use skills-based codes and brief comments rather than grades in Key Stage 4 English. The approach was devised by staff at Meols Cop High School in Southport. The FLASH Marking codes are aligned with language from the GCSE English Language and English Literature grade descriptions and are presented to pupils in order to signal where they have achieved certain skills and where there are areas for improvement and development. An example of the codes can be seen in Figure 1, which shows a small excerpt from the FLASH Marking code sheet, described in further detail in the section below.

Figure 1: Excerpt from original FLASH Marking code sheet

Why: FLASH Marking was devised with a view to improving attainment in English at Key Stage 4. The code-based approach focusing on skills, is designed to provide personalised and focused feedback to support pupil learning. The developers aimed to reduce teachers' marking workload by using this code-based approach.

Who (participants): The intervention was trialled with Key Stage 4 (Years 10 and 11) pupils working towards English Language and English Literature GCSEs. English teachers were trained in the FLASH Marking foundations and processes, and implemented these with their GCSE groups.

What (procedures, activities, and resources): Two English teachers from each 'intervention' school attended three training sessions run by the Meols Cop FLASH Marking development team. These staff were then responsible for cascading the training to all department staff who would be working with Year 10 pupils in the following academic year. The school subject leader for English was one of the staff members responsible for attending the training and they were asked to select one other English teacher to attend too. Information was provided by the delivery team to support with

the cascading, including training materials and resources, which could be shared (see below). Following cascading of the FLASH Marking training, either in June–July or September 2018, all staff with relevant GCSE English classes were expected to use the intervention with their pupils. This included sharing FLASH codes (skills) as part of lesson/learning objectives, using FLASH Marking when providing feedback on pupils' written work, tracking pupil progress using FLASH codes, and including FLASH within short- and medium-term planning.

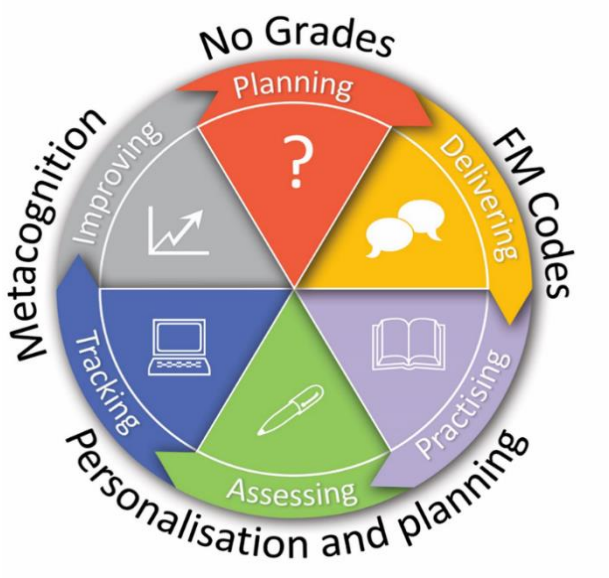
Details of the three training sessions are below. As noted above, two colleagues from each school were required to attend. In the majority of cases this did happen although there were some instances where only one teacher attended due to illness or staffing issues at school. This information was reported to us from the delivery team. Further details on numbers of teachers attending the training sessions can be seen in the IPE Results section below (see Table 16). Where teachers left the intervention school prior to Training Session 2 or 3, another English teacher was nominated to attend the session instead:

- Training Session 1 ran for a full day in June 2018. It focused on the principles of FLASH Marking, lesson planning, modelling the assessment approach, using demonstration videos, and the use of the web portal as a support mechanism for trial participants. Teachers were also informed about how to cascade the training they had received to the rest of their department. They were encouraged to do this using the materials that had been shared during the training session (for example, the FLASH code sheets, relevant PowerPoint slides).
- Training Session 2 was a three-hour session in November 2018. This session focused on moderation of assessments and the use of the codes in future planning. Support was offered through the use of demonstration videos and group discussion.
- Training Session 3 ran in July 2019, at the end of the first year of the trial. This session ran for approximately four hours and focused on reviewing progress with FLASH Marking so far, and planning for further use of the intervention in the following academic year. Teachers who attended were asked to share examples of 'good practice' with FLASH in their schools and the development team also highlighted positive examples from schools in other hubs. A range of topics were discussed in this session including, using FLASH to promote exam performance, embedding metacognition skills via FLASH, and using FLASH to reduce teacher workload.

A range of resources were developed by the development team and shared with intervention schools for use with staff and/or pupils. These included:

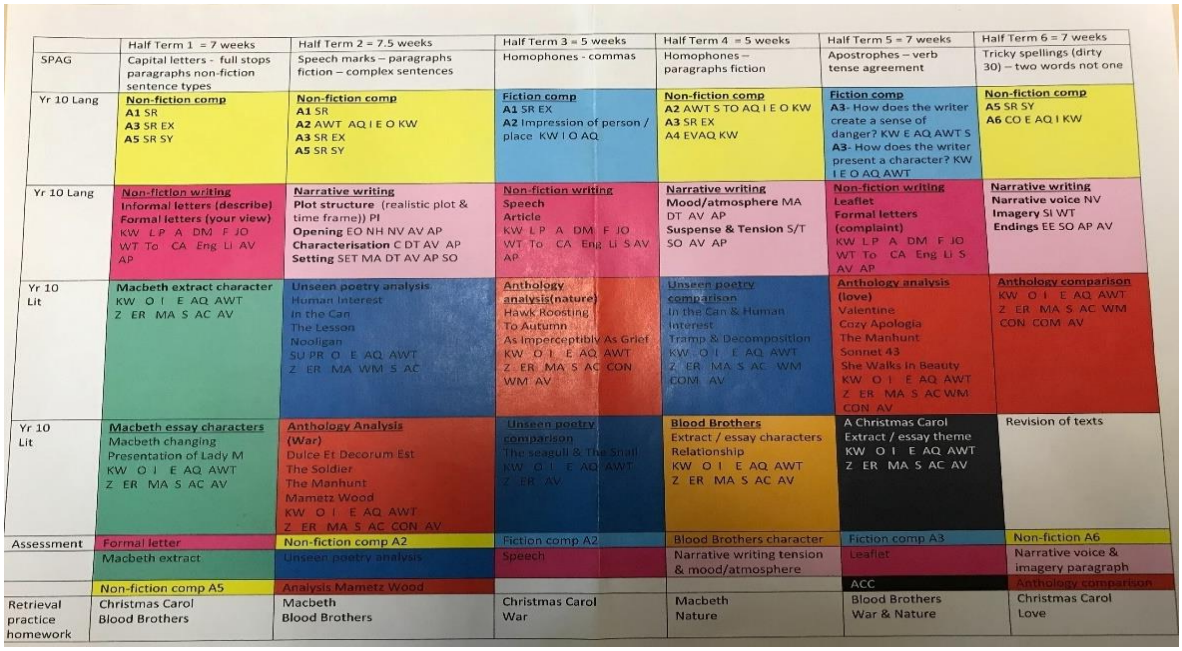
- The *FLASH Marking 'cycle' of feedback*: this visual image provided an overview of the key aims and elements of the intervention, aligned with the logic model/implementation plan (see Figure 2). This was used as a framework for understanding the different elements of the intervention, and for organising and delivering key sections of training to teachers.

Figure 2: FLASH Marking (FM) 'cycle' of feedback



- *Examples of FLASH code sheets:* the development team provided a table of codes to all schools in the first training session (see Figure 1). By the second training session, it had become clear that some schools wished to adapt their codes to align with their exam board or to include the type/number of codes, which they felt were appropriate for their pupils. The developers deemed this an acceptable adaptation as part of the intervention, and, as a result, they collated a number of different code sheets that schools were using and shared examples with teachers during the training session in autumn 2018.
- *Short-/medium-term planning documents and examples of lesson resources*: These included examples of lesson plans where FLASH codes had been included as part of the learning objectives or activities, and a scheme of work, which covered all of the Year 10 curriculum topics and their corresponding FLASH codes (see Figure 3). The aim of these plans was to encourage teachers to embed the intervention into their practice, and to consider, which FLASH codes (or skills) could be addressed best through different topics. These resources provided stimulation for discussions at the training session and ideas for planning, which teachers were asked to action when back in school.

Figure 3: Exemplar medium-term plan shared with all teachers at Training Session 2



	Half Term 1 = 7 weeks	Half Term 2 = 7.5 weeks	Half Term 3 = 5 weeks	Half Term 4 = 5 weeks	Half Term 5 = 7 weeks	Half Term 6 = 7 weeks
SPAG	Capital letters - full stops paragraphs non-fiction sentence types	Speech marks – paragraphs fiction – complex sentences	Homophones - commas	Homophones – paragraphs fiction	Apostrophes – verb tense agreement	Tricky spellings (dirty 30) – two words not one
Yr 10 Lang	**Non-fiction comp** **A1** SR **A3** SR EX **A5** SR SY	**Non-fiction comp** **A1** SR **A2** AWT AQ I E O KW **A3** SR EX **A5** SR SY	**Fiction comp** **A1** SR EX **A2** Impression of person / place KW I O AQ	**Non-fiction comp** **A2** AWT S TO AQ I E O KW **A3** SR EX A4 EVAQ KW	**Fiction comp** **A3**- How does the writer create a sense of danger? KW E AQ AWT S **A3**- How does the writer present a character? KW I E O AQ AWT	**Non-fiction comp** **A5** SR SY **A6** CO E AQ I KW
Yr 10 Lang	**Non-fiction writing** **Informal letters (describe)** **Formal letters (your view)** KW LP A DM F JO WT To CA Eng Li AV AP	**Narrative writing** **Plot structure** (realistic plot & time frame)) PI **Opening** EO NH NV AV AP **Characterisation** C DT AV AP **Setting** SET MA DT AV AP SO	**Non-fiction writing** **Speech** **Article** KW LP A DM F JO WT To CA Eng Li S AV AP	**Narrative writing** **Mood/atmosphere** MA DT AV AP **Suspense & Tension** S/T SO AV AP	**Non-fiction writing** **Leaflet** **Formal letters (complaint)** KW LP A DM F JO WT To CA Eng Li S AV AP	**Narrative writing** **Narrative voice** NV **Imagery** SI WT **Endings** EE SO AP AV
Yr 10 Lit	**Macbeth extract character** KW O I E AQ AWT Z ER MA S AC AV	**Unseen poetry analysis** Human Interest In the Can The Lesson Nooligan SU PR O E AQ AWT Z ER MA WM S AC	**Anthology analysis(nature)** Hawk Roosting To Autumn As Imperceptibly As Grief KW O I E AQ AWT Z ER MA S AC CON WM AV	**Unseen poetry comparison** In the Can & Human Interest Tramp & Decomposition KW O I E AQ AWT Z ER MA S AC WM COM AV	**Anthology analysis (love)** Valentine Cozy Apologia The Manhunt Sonnet 43 She Walks In Beauty KW O I E AQ AWT Z ER MA S AC WM CON AV	**Anthology comparison** KW O I E AQ AWT Z ER MA S AC WM CON COM AV
Yr 10 Lit	**Macbeth essay characters** Macbeth changing Presentation of Lady M KW O I E AQ AWT Z ER MA S AC AV	**Anthology Analysis (War)** Dulce Et Decorum Est The Soldier The Manhunt Mametz Wood KW O I E AQ AWT Z ER MA S AC CON AV	**Unseen poetry comparison** The seagull & The Snail KW O I E AQ AWT Z ER AV	**Blood Brothers** Extract / essay characters Relationship KW O I E AQ AWT Z ER MA S AC AV	**A Christmas Carol** Extract / essay theme KW O I E AQ AWT Z ER MA S AC AV	Revision of texts
Assessment	Formal letter	Non-fiction comp A2	Fiction comp A2	Blood Brothers character	Fiction comp A3	Non-fiction A6
	Macbeth extract	Unseen poetry analysis	Speech	Narrative writing tension & mood/atmosphere	Leaflet	Narrative voice & imagery paragraph
	Non-fiction comp A5	Analysis Mametz Wood			ACC	Anthology comparison
Retrieval practice homework	Christmas Carol Blood Brothers	Macbeth Blood Brothers	Christmas Carol War	Macbeth Nature	Blood Brothers War & Nature	Christmas Carol Love

Additional support was made available to English departments throughout the period of the trial. This was provided by the development team and included: frequent contact between schools and the development team via telephone and email; observational and support visits from the development team for the purposes of fidelity monitoring. These visits were tailored to the needs of the schools (as identified by the Heads of Department/FLASH Marking Leads and the development team) and included activities such as training and lesson observations, discussions regarding planning, resources and staff engagement, and the delivery of 'refresher' training for new colleagues, and the use of a web portal (Trello) to share videos, models of assessed work, and curriculum resources. It also involved regular communication between the subject leader and development team, and the whole English department or specific teaching staff, depending on what the school's needs were. As part of their approach to ensuring fidelity, the development team aimed to visit every intervention school at least once per year during the trial. The logistics and practicalities of organising and carrying out these visits—from both a school and development team perspective—was sometimes difficult, however. Recognising this, the development team sought to visit each school once across the trial period. A total of 32/52 intervention schools received visits prior to March 2020 and a further five schools had visits scheduled but cancelled due to the Covid-19 pandemic.

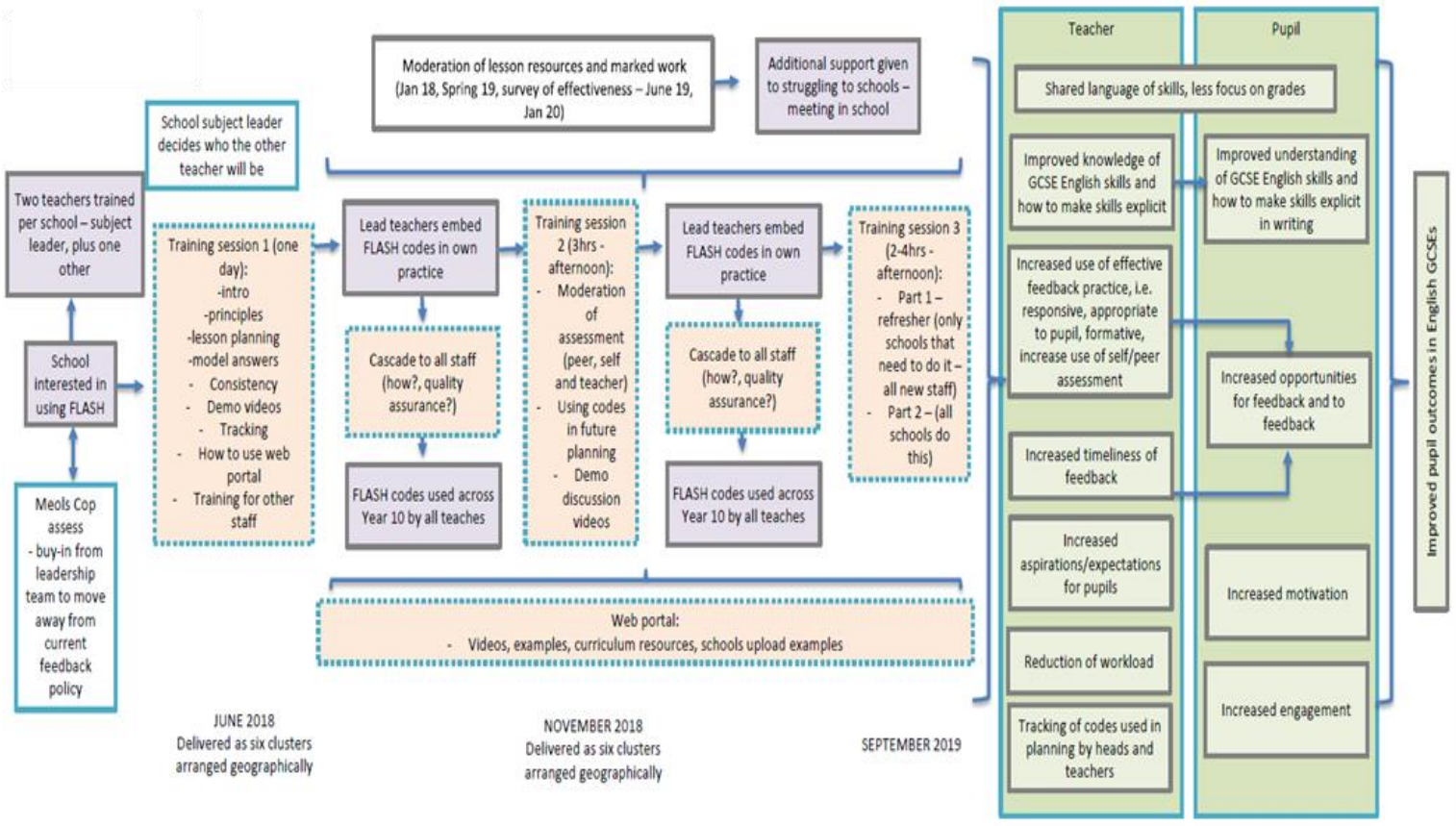
Where: Following randomisation, the 52 intervention group schools were allocated to eight training 'hubs'. These included two London and South-East Hubs plus one each based in the following areas or regions: North-West

(Liverpool); North-West (Manchester); North-West (Lake District); North-West and East Midlands (Sheffield); West Midlands; and South-West. The hub model enabled groups of schools from the same area/region to receive the training as a group and to share experiences of using FLASH Marking as part of those training sessions. The sessions took place in hotels or conference centres, which were accessible to school staff from each hub. The hubs were used for practical purposes, ensuring that the training could be delivered locally for schools taking part. They were not part of any stratification within the trial design although were used to facilitate geographical spread in the IPE. Each hub was given an author name to be identified by (due to the English Literature focus of the intervention).

Variation in delivery and implementation: A key element of the FLASH Marking approach is the opportunity for schools to develop and use the marking codes in a way that suits the colleagues and cohorts of participating schools. As such, the development team were supportive of schools adapting the codes depending upon the exam boards that they worked with, the literary texts and topics being studied, and the existing marking/feedback policies in schools. Similarly, the bespoke, ad hoc support provided by the development team was accessed and used by schools in different ways. This flexibility and variation were permitted as part of the intervention and is discussed further in the IPE Results section below.

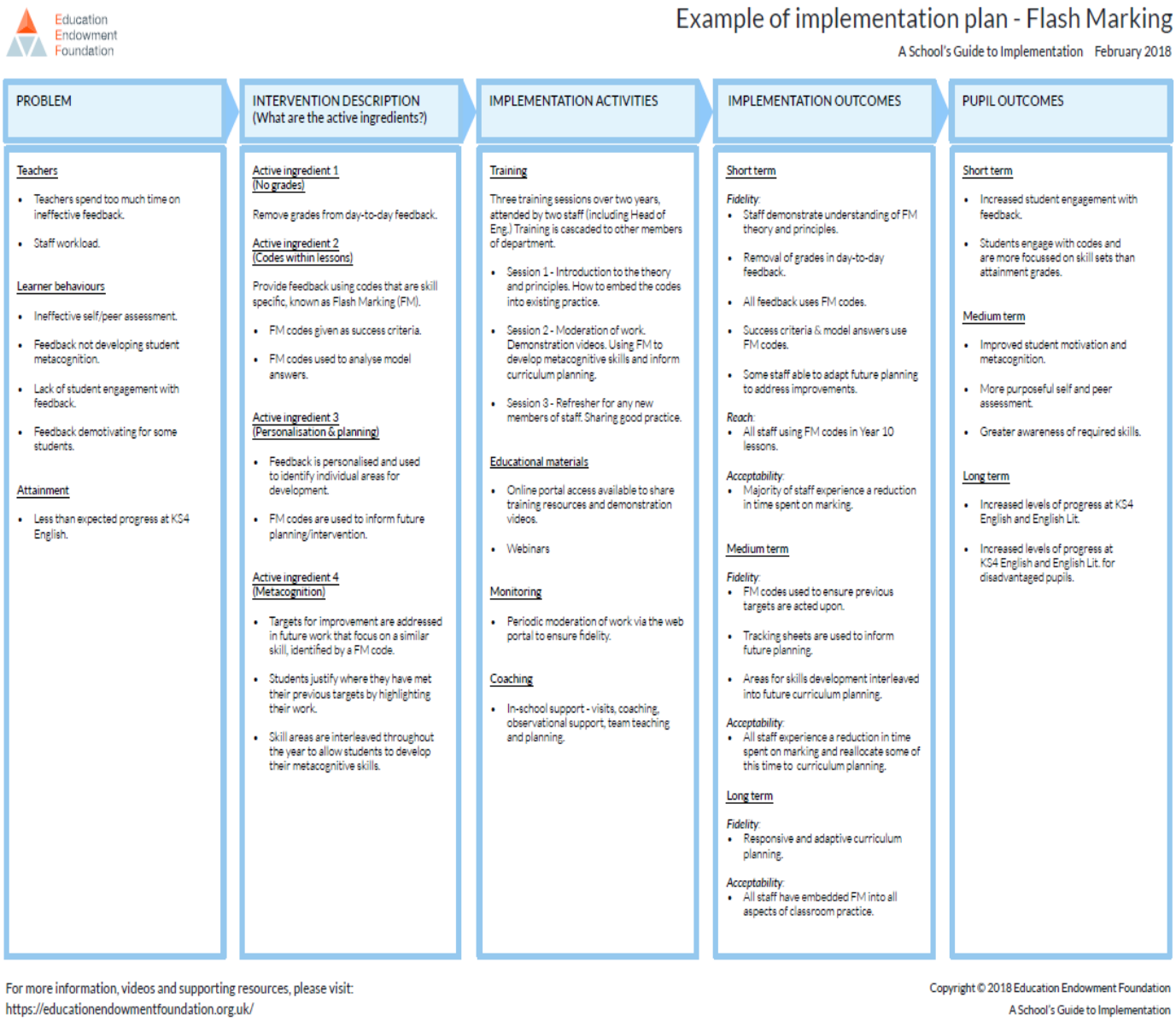
The logic model for the FLASH Marking trial, indicating the key elements and 'ingredients' of the programme and the evaluation can be seen in Figure 4. The logic model was developed at the start of the trial as a way of representing the theory of change for the FLASH Marking intervention. During initial set-up meetings, the development team and colleagues from the Education Endowment Foundation (EEF) discussed the key areas of inputs/resources relating to the intervention, its implementation, and the potential outcomes/impact; these were then organised into the logic model below. The arrows within the diagram indicate the sequence of activities over the course of the trial, and the connections, which (potentially) occur between inputs and effects or outputs. The evaluation team used the logic model, in addition to further in-depth conversations about the intervention and its implementation, to support with the design of the trial. Figure 4 shows the pre-evaluation logic model. This was updated and further summarised in the implementation plan, which can be seen in Figure 5.

Figure 4: Pre-evaluation logic model—FLASH Marking (devised by the EEF and the FLASH Marking development team during initial set-up meetings; 2017)



This logic model also informed the implementation plan below, created by the FLASH Marking development team. This was included as a model in the EEF's guide to implementation for schools (EEF, 2018), and is helpful for understanding the aims of the programme and the mechanisms (or 'active ingredients'), which might influence the outcomes of the evaluation.

Figure 5: FLASH Marking implementation plan



Evaluation objectives

The primary research question for this evaluation was:

- How effective is FLASH Marking in improving the GCSE English outcomes for Key Stage 4 pupils?

GCSE exams were cancelled in 2020 as a result of the disruptions caused by the Covid-19 pandemic. Therefore, no externally assessed GCSE data was available to measure the impact of the intervention on pupil attainment. The funders determined that it would not be possible to respond to this primary research question.

The evaluation also included a second research question, determined at the outset of the study and prior to the pandemic. Using the pre-/post-teacher workload data, we will provide a response to this question:

- How effective is FLASH Marking in reducing the marking and assessment workload for teachers of Key Stage 4 English?

As outlined in the Statistical Analysis Plan (SAP), the main analyses to respond to this question are based upon the matched pre- and post-teacher survey data. We also ran analysis on responses from all participants (for example, those teachers who had completed either one or both of the surveys) in order to report overall differences between the intervention and control groups and the two time stages.

Links to the protocol and SAP, including details of the evaluation objectives and agreed changes after the outbreak of the Covid-19 pandemic can be found on the **EEF website here.**

Ethics and trial registration

The FLASH Marking evaluation received full ethical approval from Durham University School of Education Ethics Committee.

Schools were invited to participate in the trial via email and telephone contact from the developers at Meols Cop Research School between autumn 2017 and spring 2018. A Memorandum of Understanding (MOU) was shared with schools, outlining the responsibilities of each party within the project (the school, the developers, and the evaluation team). Schools were required to sign the MOU and return it to the developers and evaluators as a condition of participation in the trial and randomisation. A copy of the MOU can be found in the appendices (see Appendix C).

Following agreement from the schools to participate, information letters and 'opt-out' consent forms were sent to all parents/carers of the Year 10 pupils to be involved within the evaluation. Information was provided about the project, including details of the intervention, the processes and methods involved for data collection, and the contact details of the development team and evaluators. Parents were asked to return a slip on the information letter if they did not wish for their child's data to be included within the project. A copy of the parent information sheet can be found in the appendices (see Appendix D).

The trial is not registered with the International Standard Randomised Controlled Trial Number (ISRCTN). At the beginning of the evaluation, it was determined that registration was not necessary due to full publication of the trial protocol and SAP and a commitment to full and transparent publication of all findings relating to the trial.

Data protection

This project was set up in 2017, prior to the introduction of the General Data Protection Regulation (GDPR). At this time, plans for data collection and data processing adhered to the 1998 Data Protection Act (DPA). For fair processing of personal data, the project adhered to a number of conditions from Schedule 2 of the DPA, including Condition 1: the data subject giving their consent (via an opt-out consent letter), and Condition 5: processing is required to carry out public functions. Prior to randomisation and following the distribution of parental consent letters, schools shared a list of Year 10 pupils' names and unique pupil numbers (UPNs) with the evaluation team. Teacher data, from the teacher workload survey, required explicit consent from those participating and was also shared directly with the evaluation team on completion of the survey.

Following the introduction of the GDPR in 2018, the evaluation team (as data controllers of the data) took a number of measures to ensure that the project was compliant with the GDPR (ICO, no date). Durham University and the University of Warwick were responsible for collecting and processing the data, were the organisations in control of personal data throughout the project, and were responsible for the production of this report. The legal basis for processing pupils' personal data for this research project is found under Article 6(1)(e) where the processing is necessary for you to perform a task in the public interest or for your official functions, and the task or function has a clear basis in law. This basis was selected as the research being carried out here is in the public interest and is being conducted by public institutions (universities). As academic organisations, a core part of our role involves engaging with and carrying out high-quality education research in the public interest (such as the project described here). We argue that participation in such activities is necessary for providing original insights in to key educational and social issues, for the promotion of evidence-informed education in England and beyond, and for educational improvement. In order to achieve these potential societal benefits, participants are needed for involvement in research studies, and their personal data is required to complete relevant analyses. The decision was made to collect the minimum amount of data possible from

schools (i.e. just names and UPNs) to reduce the amount of identifiable information that needed to be held by the evaluators. These data were then submitted to the Department for Education's (DfE's) National Pupil Database (NPD) for anonymised matching to additional information (including prior attainment and demographic variables). This baseline information would then have been used alongside post-intervention information (GCSE attainment) to examine the effect of FLASH Marking. Due to GCSE exams being cancelled, however, this information was not collected or processed.

For staff involved in the project (as data controllers), the legal basis for processing personal data is found under Article 6(1)(a) where the individual has given clear consent for you to process their personal data for a specific purpose. Teachers were required to consent to their participation in both teacher workload surveys. Information was provided to teachers prior to their engagement with the survey, and they were asked to register their consent before they could begin to complete the survey. Submitted data was stored securely on computers managed by the University Information Technology (IT) Services and protected by network firewalls and antivirus software. Submitted survey could be viewed by just two colleagues on the evaluation team. Privacy notices were drawn up and shared with the schools involved in the project (Appendix E). The schools agreed to share these privacy notices with parents and teachers, and they were also made available via the Durham University project website. These notices stated the legal basis for data processing, details of what data would be used, and contact details should participants wish to ask questions or withdraw their data. In line with the GDPR, the evaluation team also ensured that all personal data was saved on secure servers in the UK.

All data will be deleted by the evaluation team one year after the publication of the final project report. Arrangements were in place to share relevant data with the EEF archive following completion of the project. However, due to the Covid-19 pandemic, the planned analyses in relation to pupil attainment has not been possible and thus data will not be submitted to the archives.

Project team

FLASH Marking development team: Meols Cop Research School, Southport

A team from Meols Cop Research School were responsible for the development of the FLASH Marking evaluation. Leon Walker and Sarah Cunliffe, school leaders from Meols Cop Research School, developed FLASH Marking for use within their own school and piloted it with a number of local secondary schools in their area prior to the EEF evaluation. Once the trial had been agreed, Leon and Sarah took responsibility for the following activities:

- school recruitment and collection of baseline data;
- collecting opt-out consent from parents;
- ongoing communication with schools about the trial;
- training of school staff;
- delivery and development of the intervention;
- supporting cascade training in schools; and
- monitoring of FLASH implementation.

Sarah and Leon were supported by the Meols Cop Research School administrator. The development team also worked closely and communicated regularly with the evaluation team (see below) about each aspect of the project.

Evaluation team

The evaluation team was led by Dr Rebecca Morris. Prof Stephen Gorard and Prof Beng Huat See were responsible for the trial design and analysis of impact evaluation data. They also contributed to all aspects of data collection and IPE-related fieldwork. Dr Nadia Siddiqui assisted with fieldwork and IPE analysis, along with a small team of research assistants: Sophia Abdi; Dr Jack Reynolds; Dr Szilvia Schmitsek; and Lindsey Wardle.

Methods

Trial design

The FLASH Marking evaluation was designed as a two-arm randomised controlled trial (RCT) with randomisation at school level, focusing on the English GCSE outcomes of Year 10 pupils at the start of the trial (2018–2019 academic year). A total of 103 secondary schools from across England were randomised to either an intervention group (receiving FLASH Marking) and a 'business as usual' group (continuing with their usual assessment and feedback practices). All English departments within schools allocated to the intervention group were eligible for the FLASH Marking training and support from the development team. All Year 10 pupils within the participating schools during the academic year 2018–2019 were eligible to participate in the trial. Parents were given the opportunity to withdraw their child's data from the evaluation. Pupils in the treatment schools still received the intervention (as part of their English lessons) but none of their data was collected or processed as part of the evaluation.

The Covid-19 pandemic and its impact on the FLASH Marking trial

As we note above, the FLASH Marking evaluation was significantly affected by the Covid-19 pandemic in 2020–2021. The original trial was due to use GCSE data from 2020 for the trial cohort (pupils in Year 10 at the start of the trial) in order to establish the impact of FLASH on attainment in English. When externally awarded GCSEs were cancelled during the first year of the Covid-19 pandemic, it became clear that conducting this analysis would not be possible. The evaluation team and the EEF explored options to complete the trial with an attainment outcome. During autumn 2020, plans were developed to conduct impact analyses using GCSE attainment data for pupils who had been in Year 9 at the start of the trial and in Year 11 in 2021. At this point, GCSE examinations had been expected to revert to normal and to be awarded externally rather than by teachers/centres. It was deemed appropriate to consider this amended approach for analysis as the teachers in the treatment schools had been engaged with the FLASH intervention and many schools were using the approach with year groups other than the original trial cohort. All schools in the trial (both intervention and control) were contacted regarding a possible extension to the evaluation and with a view to beginning a process of collecting relevant data from the new cohort of pupils. In January 2021, however, the government announced that externally awarded GCSEs would be cancelled for a second year. At this point, plans for the trial extension were cancelled and the decision was made to complete the analysis and reporting of the evaluation based upon the secondary outcome data (teacher workload) and findings from the IPE.

Data from both the teacher workload questionnaire and the IPE approaches had mostly been collected prior to the Covid-19 school closures. As such, the evaluation team felt that there were enough relevant data here to report on. As we note below however, the second teacher questionnaire was available to staff to complete during late-February 2020–March 2020. This was at a time when challenges relating to Covid-19 were becoming more apparent and were requiring teachers' time and input. As a result, it is likely that pressures caused by the Covid-19 pandemic impacted the questionnaire return rate.

The evaluation team had been collecting IPE data since the very beginning of the trial in summer 2018. We had visited many of our case study schools three times out of the four that we had originally intended and had extensive data from other sources too (training observations, questionnaires, and correspondence between schools and the development team). Due to the challenges and pressures that teachers were facing during the pandemic in spring/summer 2020, we decided not to engage them in further, virtual data collection activities (for example, online interviews). The final IPE school visit took place in early March 2020 and the online questionnaire was closed at the end of March 2020. We recognise that this may have resulted in the potential loss of some relevant IPE information; however, given the unprecedented circumstances of the time—and the amount of data already collected, we argue that our existing dataset is sufficient for understanding the implementation of FLASH Marking in schools and its potential to be scaled-up in future.

Table 2: Trial design with modifications due to the Covid-19 pandemic

Trial design, including number of arms		Two-arm, cluster randomised controlled trial
Unit of randomisation		School
Stratification variable(s) (if applicable)		N/A
Primary outcome	Variable	N/A (no primary outcome due to cancellation of GCSE exams)
	Measure (instrument, scale, source)	N/A
Secondary outcome(s)	Variable(s)	Reported teacher workload
	Measure(s) (instrument, scale, source)	Teacher workload survey (number of hours spent on teaching and learning activities per week)
Baseline for primary outcome	Variable	N/A
	Measure (instrument, scale, source)	N/A
Baseline for secondary outcome(s)	Variable	Reported teacher workload (pre-intervention)
	Measure (instrument, scale, source)	Teacher workload survey (number of hours spent on teaching and learning activities per week)

Participant selection

Schools

All mainstream, state-funded secondary schools in England were eligible for inclusion within the FLASH Marking evaluation. As per the MOU (Appendix C), schools were required to confirm that they were: willing to be randomised to either the intervention or control groups; willing to participate in the FLASH Marking intervention if allocated to the treatment condition; willing to support the evaluation through the provision of data for the impact and process evaluations; and communicate with the evaluation team as needed. Prior to randomisation, schools were required to complete and sign the MOU, share baseline pupil data, and ask English teachers to complete the pre-intervention workload survey.

The delivery team was responsible for recruitment of schools. At the start of the project, they attempted to share details of the FLASH Marking project and an invitation to participate (via email) with every mainstream secondary school in England (c.3,400 schools). Approximately 600 schools expressed an interest in participating in the evaluation and 269 schools completed the initial paperwork (an expression of interest form and provision of staff contact details) and were assessed for their eligibility. For inclusion at this stage, schools were required to be mainstream secondary schools in England and they were asked to confirm commitment to the key elements of the study. Schools with a higher than (national) average proportion of socioeconomically disadvantaged students were also encouraged to apply, in line with the EEF's commitment to improving outcomes for students from less affluent backgrounds. Schools were excluded at this stage if they were not situated in England, if they were a primary or middle school (without a Key Stage 4 cohort), and if they were independent/private fee-paying schools. Following this first stage of screening, eligible schools were invited to complete the MOU, share information with Year 10 parents about the project, provide pupil information, and share the first teacher workload survey with English departments. Those completing these steps were recruited to the trial on a 'first come, first served' basis.

The project had the capacity for 100 schools to participate and this was eventually extended to 110 in order to accommodate more schools and to meet the seemingly high demand for involvement. In January 2018, a total of 110 schools who had completed all required paperwork and were eligible for inclusion, were confirmed as participants; seven of these schools pulled out of the trial or were excluded at this stage. Withdrawal was due to unforeseen staffing or school-level changes, and one school was excluded due to being an English-medium school in India. A total of 103 schools were identified for randomisation. The delivery team passed these details to the evaluation team who carried out simple randomisation and informed the delivery team of the outcomes to be shared with schools.

Incentives were offered to participating schools to support recruitment and to encourage engagement with the intervention and the evaluation. Schools in the intervention arm were given £700 for participation; those in the business as usual group were given £1,000. Payment of these incentives was dependent upon meeting certain conditions such as providing baseline data, attending training, and supporting evaluation activities.

Table 3 provides an overview of the characteristics of the 103 schools, which were randomised as part of the FLASH Marking trial. Generally, the schools are balanced in terms of these characteristics. The cohorts have the same average Key Stage 2 points score, for example, and the school sizes are similar. Pupils in the intervention group are slightly less likely than those in the control group to be eligible for free school meals (FSM) and slightly less likely to have English as an Additional Language (EAL). There were also a higher proportion of 'Outstanding' schools in the intervention group (n=16 schools versus n=12 in the control group); however, if combined with the number of 'Good' schools, the proportions are similar (78.9% and 76.4%, respectively). Schools in the intervention group were more likely to be academies and free schools (75%) than those in the control group (58.5%).

Table 3: School characteristics: baseline characteristics of school groups as randomised

School level (categorical)	**Intervention group (n=52 schools)**		**Control group (n=51 schools)**	
	N (missing)	**Percentage**	**N (missing)**	**Percentage**
Ofsted rating				
Outstanding	16	30.8	12	23.5
Good	25	48.1	27	52.9
Requires improvement	7	13.5	7	13.7
Inadequate	2	3.8	2	3.9
No report	2	3.8	3	5.9
School type				
Academy convertor	26	50.0	20	39.0
Academy sponsored	11	21.2	9	17.6
Community school	7	13.5	12	23.5
Foundation school	1	1.9	6	11.8
Free school	2	3.8	1	1.9
Voluntary aided school	4	7.7	3	5.9
Voluntary controlled school	1	1.9	0	0
Location: urban/rural				
Urban	43	82.7	44	86.3
Rural	9	17.3	7	13.7
School level (continuous)	**N (missing)**	**Average (national average)**	**N (missing)**	**Average (national average)**
School size	52 (0)	1,003 (948.3)	51 (0)	1,034 (948.3)
Proportion pupils eligible for FSM	52 (0)	12.8 (13.3)	51 (0)	14.2 (13.3)
Proportion pupils with EAL	52 (0)	11.9 (16.6)	51 (0)	14.3 (16.6)
Key Stage 2 average points score	52 (0)	28.3 (25.9)	51 (0)	28.3 (25.9)

Data sources: DfE Annual Schools Census and Performance Tables (DfE, 2018a; 2018b) and Ofsted website (Ofsted, current).

Pupils and teachers

FLASH Marking is an English subject-based intervention. Therefore, within participating schools, it was English departments and specifically English teachers with Key Stage 4 classes who were involved within the project. After randomisation, intervention schools were asked to select two English teachers (one Head of Department and one additional teacher) to attend the training. All English teachers teaching Key Stage 4 classes were expected to deliver the intervention following cascading of the training.

All Year 10 pupils (in the academic year 2018–2019) in participating schools were eligible for inclusion within the evaluation. They would be included in the trial for the duration of the intervention, for instance the two years of their Key

Stage 4 GCSE course. A total of 18,521 pupils and 990 teachers (472 in the control group and 518 in the intervention group) were included in FLASH Marking at the start of the trial. A small number of pupils (n<10) had their data withdrawn by parents prior to the start of the trial but all continued to engage with either the intervention or control conditions as part of their English GCSE curriculum.

Outcome measures

Primary outcome

No primary outcome was collected due to the cancellation of externally assessed GCSEs as a result of the Covid-19 pandemic (see Trial Design section).

Further details of the intended primary outcome measures can be found in the original **evaluation protocol** and **SAP**.

Secondary outcomes

The secondary outcome for the evaluation focuses on reported teacher workload. This outcome was identified as a potential benefit of FLASH Marking during the early piloting stages of the intervention. In line with the logic model, the development team felt that FLASH could reduce teacher workload due to the decrease in the amount of time required for writing extensive comments regularly on pupils' work.

English teachers with Key Stage 4 classes in all participating schools were asked to complete a pre- and post-intervention online survey. Both surveys were designed by the evaluators and shared with trial schools by the development team. The first survey included a total of 20 items, and asked teachers to share the number of hours that they spent on different teaching duties (for example, feedback and marking, planning lessons, and communicating with parents) and their total hours on teaching and learning activities in their most recent full working week. Participant teachers were also asked to report on their school's marking and feedback policies, the type and amount of marking carried out with Key Stage 4 classes, and their views on their workload. The sections asking about reported hours on different activities and on views of workload were taken from a recent DfE instrument (DfE, 2016) used as part of the government's Workload Challenge project. The aim of this survey was to establish teachers' working practices (in relation to marking/feedback and other activities) at the outset of the project. The first survey was piloted at the development team's school and then shared with project schools in January 2018. Completion by Key Stage 4 English teachers in trial schools was a condition of randomisation.

The second survey was administered to schools in February 2020, towards the end of the FLASH Marking intervention. This survey was piloted in one intervention school in January 2020 prior to it being shared with all evaluation schools. As with the first instrument, the survey was distributed by the development team to the English teacher contacts that they held. The survey included the same items as the pre-intervention instrument (in order to provide pre-/post-comparisons between teachers' reported workload and attitudes towards their workload). It also included two additional questions about experiences of using FLASH Marking for teachers based in intervention schools, and a further two questions about department use of the intervention for English Heads of Departments. These latter questions for intervention school staff were aligned with process evaluation outcomes where we were examining staff attitudes to the intervention. Further, Heads of Departments were asked to report on their department's commitment to FLASH as one of the compliance measures (see below). Due to the Covid-19 pandemic and the challenges this presented for school teaching staff, plus other factors such as staff turnover, the return rate for this survey was lower than originally anticipated. We discuss this further below.

Copies of both survey instruments can be found in Appendices E and F.

Sample size

Prior to recruitment, the intention was to recruit c.100 schools for this trial. This was then increased to 110 following considerable interest in the project from schools. A total of 103 schools were randomised: 52 to the intervention group; and 51 to the control group.

The evaluation team calculated the sample size required for an 'effect' size to be considered secure. These are not reported here as they relate to the primary outcome (pupils' GCSE attainment) but see **original protocol** and the **SAP** for further details.

Randomisation

Randomisation was conducted by the evaluation team and took place in spring 2018 following recruitment of 103 schools to the project. As per the protocol, a simple randomisation process was used with an online randomisation programme (randomiser.org). Randomisation to two groups took place before a final random decision as to which group would receive the intervention. As 103 schools were eligible for randomisation, a decision was taken over how to allocate the 103rd school (due to there being an odd number). Due to the enthusiasm for the FLASH Marking intervention from those who had signed up to participate in the trial, the evaluation team decided to allocate the 103rd school to the intervention group. This resulted in 52 schools being allocated to the intervention group and 51 to the control group (see the SAP for the CONSORT diagram describing numbers of schools engaged at each point of the trial). Over 9,000 pupils were allocated to each arm of the trial.

A national sampling frame was used and all 103 schools were randomised in a single batch. There was no stratification by region. The eight regional hubs were introduced for training and convenience purposes and were determined after randomisation.

Statistical analysis

Primary analysis
As noted above and in the updated protocol and SAP, no primary analysis was possible due to the lack of externally assessed GCSE scores in 2020.

Details of the planned approaches for analysis can be found in the **original versions of the protocol** and the **SAP** on the EEF website.

Secondary analysis

For secondary outcome analysis, we focused on estimating the effect of the intervention on teachers' workload. A baseline workload survey was completed (before randomisation) by Key Stage 4 English teaching staff in trial schools. Items from the survey that will be included in the secondary analysis are: (1) items asking teachers to report the number of hours that they spend on different activities; and (2) items relating to the total number of hours worked for their last full working week (see copy of survey in Appendix F). They were also asked to report perceptions of their workload. A second survey (Appendix G) was administered during the early spring term of the second year of the trial (January 2020–February 2020) with a view to examining whether time spent on marking/feedback has altered for those teachers within FLASH Marking schools.

The analysis focuses on the differences between teachers in the control and intervention groups at the two measurement points (first survey and second survey). 'Effect' sizes are used to examine the differences between the two groups with regard to the number of hours that teachers reported spending on teaching activities overall and the amount of time spent on different aspects of their job (including assessment and marking). The response rate for the second survey was considerably lower than for the first. We speculate that this could be for a number of reasons, including staff turnover, Covid-19-related challenges, and teachers' interest or commitment to FLASH Marking; the response rate to the second survey was considerably lower than for the first. We summarise the level of missing data, and these missing cases are examined in order to establish whether there are differences in the pretest scores of missing cases between the two groups. Categorical variables (i.e. items about teachers' attitudes to their workload) will also be analysed using odds ratios to examine changes between pre- and post-measurements. An odds ratio is a measure of association between an exposure and an outcome. It represents the odds that an outcome will occur given a particular exposure (a/c), compared to the odds of the outcome occurring in the absence of that exposure (b/d). Therefore, the ratio of a/c to b/d is ad/bc. See Gorard (2021) for further explanation.

Estimation of effect sizes
As per the EEF current guidance, 'effect' sizes will be calculated using Hedges' g for each variable based on the difference

between mean post-test (and gain scores) for each relevant variable, such as the number of hours reported for marking and feedback. We will not report 'confidence intervals', but an interested reader can compute them if they wish as we will report the number of cases per group, standard deviations, and the effect size for each comparison.

For ease, the Hedge's g 'effect' size formula is written out as follows:

$$Effect\ size = \frac{[mean\ of\ treatment\ group] - [mean\ of\ control\ group]}{standard\ deviation\ (pooled)}$$

Confidence intervals are not reported as they are based on unwarranted assumptions, and are routinely mis-interpreted (Gorard, 2021). However, an interested reader can compute them if they wish as the number of cases per group, and the effect size for each comparison, are presented in this report.

Implementation and process evaluation

As set out in the protocol, the IPE was designed to understand a range of issues concerned with the implementation of FLASH Marking in trial schools. Following data collection and analysis, we have consolidated the core aims of the IPE into the following subheadings, including all of those noted in the protocol and SAP documents. As such the IPE addresses the following key areas:

- compliance and fidelity to the FLASH Marking intervention:
    - teachers' delivery of the intervention; and
    - possible indication of contamination or diffusion.
- factors which influenced the delivery and implementation of FLASH Marking in schools:
    - teachers' response to training, including attendance rates;
    - fidelity of training;
    - understanding of the process and purpose of FLASH Marking; and
    - the contents and use of any materials.
- perceived outcomes of the intervention and its implementation:
    - staff and pupils' views of the intervention;
    - perceived impact on children's behaviour and attitudes; and
    - changes in teachers' behaviour.

Linked to the above factors, we also include discussion of barriers to implementation that were identified and the potential for the FLASH Marking intervention to be 'rolled-out' beyond trial schools.

Our framework for the IPE was informed by the logic models developed by the Meols Cop development team and the EEF (see Figures 4 and 5). These provided clarity around the intended activities and outcomes for schools involved in the FLASH Marking trial. Given the long-term nature of the intervention (i.e. across two academic years) and the intention to examine both attainment outcomes and teacher workload, we designed an IPE, which would capture information from a range of sources and participants. The aim was to provide both a holistic and in-depth picture of the contexts in which FLASH Marking was being implemented and of the perspectives of those involved in the trial. Moreover, and in line with the logic models, the IPE was designed to provide further explication of both the 'ingredients' of FLASH Marking (i.e. the components, which form the intervention) and the mechanisms, which might lead to the desired—or other—outcomes. Better understanding of these elements and of the ways they work and interact together is important for informing the design and implementation of future interventions in this area.

IPE research methods

The IPE employed a mixed-methods approach. As part of the initial recruitment and signing-up procedure, participating schools were informed about the IPE and were asked for their cooperation and involvement, where needed, from the

outset. Table 4 provides an overview of the different methodological approaches used for collecting IPE data; we then provide some further detail on each of these strategies for data collection and analysis.

Table 4: FLASH Marking implementation and process evaluation (IPE) methods overview

Research methods	Data collection methods	Participants/ data sources	Data analysis methods	Implementation/ logic model relevance
School case studies (n=16)	Observations of lessons, interviews, examination of pupil exercise books, and other relevant documents	Staff (teachers and school leaders) and pupils	Thematic analysis	• Understanding the extent to which FLASH Marking has been implemented as intended • Understanding teachers' and pupils' experiences of FLASH Marking implementation (and the business as usual control) • Exploring teachers' and pupils' perspectives on FLASH Marking and views about perceived outcomes
Teacher questionnaire (pre- n=833 teachers; post- n=358 teachers)	Questionnaire	English teachers	Simple numerical analysis (frequencies, percentages, standard deviation (SD), and correlation)	• Understanding the extent to which FLASH Marking has been implemented as intended • Understanding teachers' experiences of and attitudes towards FLASH Marking implementation (and the business as usual control) • Exploring teachers' views about perceived outcomes
Pupil questionnaire (n=8 schools; 474 pupil participants)	Questionnaire	Year 11 pupils in intervention schools	Simple numerical analysis (frequencies, percentages, SD, and correlation)	• Understanding whether pupils are aware of FLASH Marking and use it within their English classes • Exploring pupils' perspectives on FLASH Marking and views about perceived outcomes
Observation (n=10/24 sessions)	Observation of training sessions	Developers and English teachers	Thematic analysis	• Understanding the content and delivery of the training and how this aligns with the logic model. Also, gaining insights into teachers' experiences and attitudes towards the training
Correspondence	Emails and training session evaluations (anonymised and provided by the FLASH Marking development team)	English teachers from intervention schools	Thematic analysis	• Understanding experiences and attitudes towards the FLASH Marking intervention, plus perceived outcomes as noted by participating teachers

Case studies

Data collection from a number of case study schools formed the core element of our IPE work. The aim of the case study schools was to provide detailed insights about the intervention and its implementation across the period of the project. At the outset of the study, the evaluation and development teams and the EEF determined that the identification of a number of 'case study' schools would be used as examples of the implementation of the FLASH Marking intervention, providing a sustained approach to data collection over the course of the trial and facilitating a holistic understanding of the intervention in practice. We recognise that the term 'case study' is widely used in research across the social sciences and beyond; there is no single understanding of what a 'case study' is nor how the approach should be used for collecting and collating data. With this in mind, we draw upon this inclusive definition from Thomas and Myers (2015).

> *Case studies are analyses of persons, events, decisions, periods, projects, policies, institutions or other systems which are studied holistically by one or more methods. The case that is the subject of*

> *the inquiry will be an instance of a class of phenomena that provides an analytical frame—an object—within which the study is conducted and which the case illuminates and explicates.*
>
> (Thomas and Myers, 2015, p. 7)

In this evaluation, the individual schools selected as 'case studies' provide us with a clear analytical frame for illuminating and explicating the implementation of FLASH Marking.

For each case study school, a range of data was collected, using observation, interview, and documentary analysis approaches. At the beginning of the trial, the evaluation team, with input from the developers, planned to identify 12 intervention schools and three control schools as 'case studies'. Following discussions with the development team in the early stages of the trial, the decision was taken to invite an additional intervention school to participate in case of dropout, taking the total to 16 case study schools (13 intervention, three control). By the second year of the evaluation, two of the case study schools from the intervention group had withdrawn from implementing FLASH (due to staff changes and revised school improvement priorities) and were no longer able to contribute to the IPE elements of the evaluation. Two further schools were recruited as case studies for the remaining year of the evaluation. Their details are provided in the overview in Table 5 below.

The evaluation team planned to visit each case study school four times over the course of the trial; twice in the first academic year and twice in the second. For a number of reasons (for example, staff illness, staff availability, Ofsted visits, and Covid-19 school closures), it was not possible to visit every school four times. In order to make the multiple visits to schools worthwhile, the evaluation team adapted the foci for each round of visits, responding to the different phases and elements of the trial (see logic models above). During the first visit, for example, there was a general focus on teachers' experiences of the training and cascading, and an examination of how the introduction of FLASH had been received by staff and pupils. In follow-up visits, we were aware of further training sessions that had occurred and also sought to explore how embedded FLASH was within each school.

Case study schools were selected based upon the following:

- Regional hub that they were part of. We wanted to ensure a spread across all eight hubs and so ensured that there was at least one intervention group case study school from each hub.
- Variation in school type, context, and demographic of pupils. Of course, it was not possible to gather a truly 'representative' sample of schools across the group of participating schools. However, we did want to understand how FLASH Marking worked in different settings and with different groups of pupils. As such, we looked to include schools who had different characteristics in terms of geographical location, urban/rural setting, high/mid/low levels of socioeconomic disadvantage, and high/mid/low performance information (based on GCSE attainment and Ofsted gradings).
- Willingness of the school and English department to participate as a case study school. Information about the IPE and what it would involve in terms of school visits and data collection was shared with the English Head of Department. Permission was sought from them prior to any case study data collection being undertaken. In order to encourage schools to participate, the evaluation team clearly explained the expectations required from schools and reiterated the voluntary nature of this element of the study. We sought to reduce burden on case study schools by keeping all data collection activities as informal, 'light touch', and simple as possible. All schools who were approached to be case studies in autumn 2018, agreed to this. Some later withdrew, as we discuss further below, but were replaced with alternative schools who were willing to participate.
- For the three control group, case study schools, we recruited schools from three different regions and with different pupil demographics. Again, representativeness was not the main aim here; instead, we were focused on trying to understand some of the 'business as usual' approaches that were being used in schools at the same time as FLASH Marking was happening.

Table 5 summarises information about each of the case study schools along with the number of visits that the evaluation team made to each school. Contextual information about each school has been gathered from DfE (2018a, 2018b; 2019) and Ofsted websites. It should be noted that schools K and M were the two that stopped using FLASH Marking after the first year of the trial. These schools were replaced with schools N and O as case studies.

Table 5: Case study schools: contextual information

School	School type	Urban/rural	FSM %	White British %	EAL %	Ofsted grade	Progress 8 score	Attainment 8 score	Attainment 8 (disadvantaged pupils)	No. of visits from evaluation team
Intervention group										
A	Academy sponsor led	Urban major conurbation	54	11	59	Good	0.4	48	47	3
B	Academy converter	Rural town and fringe	5	82	8	Good	-0.2	48	29	3
C	Community school	Urban major conurbation	34	89	3	Requires improvement	-0.5	40	34	3
D	Academy sponsor led	Urban major conurbation	13	73	13	Outstanding	-0.1	40	34	2
E	Academy converter	Urban major conurbation	6	48	31	Good	-0.1	48	37	3
F	Academy converter	Urban major conurbation	12	64	15	Good	0.3	49	39	3
G	Academy converter	Urban minor conurbation	10	92	3	Requires improvement	-0.2	45	35	4
H	Academy sponsor led	Urban city and town	35	94	0	Good	0.0	44	40	2
I	Academy sponsor led	Urban city and town	20	88	8	Good	0.5	51	45	3
J	Academy sponsor led	Urban major conurbation	21	16	54	Good	0.4	44	42	4
K	Academy converter	Urban city and town	15	92	3	Good	-0.7	35	28	2
L	Academy converter	Rural town and fringe	6	90	4	Good	0.4	51	46	3
M	Academy converter	Urban city and town	26	90	4	Requires improvement	-0.8	36	27	1
N	Academy sponsor led	Rural town and fringe	14	95	2	No published Ofsted grade	-0.0	41	30	1
O	Academy sponsor led	Urban city and town	26	90	7	No published Ofsted grade	-0.4	46	36	1
Control group										
P	Academy converter	Urban city and town	11	87	7	Good	0.1	46	31	3
Q	Community school	Rural hamlet and isolated dwellings	4	92	1	Good	0.4	55	45	3
R	Free school	Urban major conurbation	12	1	42	Outstanding	1.0	56	52	2

Note: information relating to the geographical region of schools has not been included in order to preserve anonymity of the participating schools, teachers, and pupils. Percentages have also been rounded to the nearest 'one' (or 0.1 for Progress 8 scores) in order to remove the possibility of identification.

Case study methods

Observations

Wherever possible, the evaluation team observed Key Stage 4 English lessons during their case study visits to schools. When setting-up the visits, Heads of English were asked to facilitate these observations, allowing us to either view whole lessons or to observe parts of lessons where this was more appropriate or convenient. Decisions about which lessons or teachers to be observed were left to the Head of English, although in our correspondence to arrange visits we were clear that we were keen to see a range of different pupils and staff. The number of lessons observed varied from school to school; in some schools we saw just one lesson, whereas in others we were invited to observe up to four different classes/teachers. Table 6 summarises how many lessons were observed in each school across the two years of the trial.

Table 6: Number of lessons observed and interviews conducted in each case study school

School	No. of visits from evaluation team	No. of lesson observations (full or part lessons)	No. of staff interviews	No. of pupils interviewed
Intervention group				
A	3	14	8	18
B	3	10	12	19
C	3	3	7	10
D	2	6	5	15
E	3	7	4	12
F	3	3	9	12
G	4	8	8	16
H	2	4	8	6
I	3	6	5	12
J	4	14	15	22
K	2	3	5	14
L	3	12	12	24
M	1	1	4	6
N	1	3	3	6
O	1	1	1	6
Control group				
P	3	14	7	8
Q	3	3	4	12
R	2	4	3	4

Prior to our visits, we made it very clear that during our observations we would not be making any judgements about the 'quality' of teaching taking place, but instead were focusing on planning and pedagogy relating to marking and feedback (and particularly FLASH Marking in the intervention schools). Observers from the evaluation team were briefed about the elements and areas that they should be focusing on during the English lessons. These included how FLASH Marking was being used, resources including FLASH codes, the way teachers talked about/used FLASH Marking, pupils' responses to FLASH Marking as part of their learning, and general assessment and feedback approaches (see Appendix H for an example of the briefing sheet provided to observers to use during school visits). In control schools, observers focused on existing marking/feedback strategies. The structure of these observations was deliberately 'open' in order to allow observers to write detailed fieldnotes about areas that they also felt were pertinent. Our aim as observers was to be as discreet as possible and avoid changing the dynamic, content, or delivery of the lessons. In some schools, teachers preferred us to sit at the back of the class to view the lesson and to take notes; in some other schools, teachers were keen for us to move around the room, look at pupils' books, and talk to them about their experiences of English and FLASH Marking. The visiting evaluators were flexible in their approach here, in order to accommodate the teachers' wishes and ensure that they felt at ease having visitors in the classroom. The teachers that we encountered were very welcoming and open to us joining their classes.

Interviews

In addition to observing English lessons, evaluators carried out interviews with staff and pupils at the case study schools (see Table 6 for the number of interviews in each school). These varied in level of formality, as determined by the availability and willingness of suitable participants. As a minimum, most visits included a discussion with the Head of Department and/or FLASH Marking Lead at the school. Sometimes these were lengthier interviews, following a sequence of questions/topics provided to the visiting evaluators as part of their briefing. In other cases, this more in-depth interview approach was not appropriate and so short discussions would be carried out instead, often with a range of relevant colleagues. Examples of this include speaking to a teacher in the playground while they were on break duty, or in the staffroom during lunch. Again, we felt that this flexibility was important in allowing us to reach a wider range of colleagues within schools and also for ensuring that they felt at ease talking to us. In order to capture an authentic description of their experiences as part of the trial, we wanted busy school leaders and teachers to feel that they could speak openly and honestly. Allowing for a more unstructured, informal approach was, at times, helpful for contributing to this.

Heads of Department or FLASH Marking Leads at case study schools would signpost us to teachers who would be appropriate and available for interview. Sometimes this also 'snowballed', and other colleagues would be included. At one school, for example, a teacher suggested we speak to two trainee teachers who were using FLASH Marking in their lessons. At another school, we were advised to speak to the school librarian who had supported the introduction of FLASH and developed a display to share the codes with pupils. Again, we were open to this approach and found that it provided us with a broader, richer set of data to understand the implementation of the intervention. On average, we engaged with four teaching staff on each FLASH Marking visit. Again, this varied from school to school, depending very much on the availability of teachers and the schedules that Heads of Department had determined for our visits.

In all case study schools the evaluation team were able to engage with Key Stage 4 pupils studying English Language and English Literature. As with teaching colleagues, facilitation of this was the responsibility of the Head of Department or FLASH Marking Lead, and schools managed this in different ways. Some schools were very happy for us to speak to pupils during lessons. Where this was the case, we would have brief discussions with pupils either in small groups or individually. In approximately half of the intervention school visits, teachers organised focus groups of pupils for us outside of lessons. These focus groups tended to happen during break or lunchtime and included small groups of between four and ten pupils; a teacher would be in the same room but did not participate in the discussions. Being able to engage with pupils in this way was important for gaining a pupil perspective. A number of key questions were put to pupils (see an example on the briefing sheet, Appendix H). These served as helpful prompts for starting discussions and for allowing pupils to expand and develop their responses. Again, a relatively informal and relaxed approach appeared to yield candid and thoughtful contributions from pupils, all of which have been helpful for informing our understanding of how FLASH was working and the perceived effect on their learning.

Consent for participation in the interviews was provided via the parental opt-out forms issued to parents at the start of the trial. We also requested verbal consent from pupils prior to any interviews/discussions taking place. To record information from interviews with staff and pupils, evaluators made detailed field notes both during the interactions and immediately afterwards. While this does risk a loss of information or inaccurate capturing of responses (as opposed to more formal digital recording and transcription of data), we deemed it the most appropriate approach for the sometimes spontaneous data collection that occurred and for putting participants at ease.

Documents/exercise books

Our case study school visits also included gathering information from other sources such as pupils' exercise books, classroom displays, and teachers' planning documents or resources. Teachers shared these with members of the evaluation team, usually during the visits but sometimes afterwards (via email too). Pupils' exercise books were particularly helpful for gaining an understanding of the extent to which FLASH Marking (or other approaches) were being utilised and also *how* these approaches were being used. We could see, for example, whether pupils had FLASH Marking code sheets in their books and where they were using coloured highlighters for reviewing their written work; we could also note where teachers were using the codes in their feedback to pupils, and what this looked like in practice. In some of the pupil focus groups that we conducted, pupils came with their exercise books. This provided an excellent stimulus for discussion around their learning in English and enabled them to refer to examples from their books.

Information from these documentary sources was recorded using a camera. Permission was always sought from the pupils and teachers prior to photographing any work, displays, or other sources. No pictures or names of children/staff were photographed. Examples from this valuable source are included in our results section below.

Questionnaires

Teacher questionnaire

To measure teachers' reported workload, an online survey was used pre- and post-intervention (see above). In the second (post) teacher survey, we also included a series of questions in order to gather information for the process evaluation (see Appendix G). These questions focused on teachers' attitudes to their workload and to the approaches used for marking/feedback in their school (for intervention schools, this included FLASH Marking). In intervention schools, teachers were asked to report on their views of FLASH Marking, its implementation and the training received via a series of Likert scale questions. Teachers were also invited to leave longer comments if they wished. Finally, Heads of Department were also asked to respond to two further closed questions about the influence of FLASH Marking on their department and whether they intend to continue using the intervention after the end of the trial. Both survey instruments were developed by the evaluators in close collaboration with the development team. They focus on teachers' experiences and attitudes towards FLASH Marking, in line with the logic model/implementation plan and the aims of the IPE (as stated above). The first survey was piloted with English teachers in the development team's school, and the second survey was piloted with a large English department from the intervention group of the trial.

Pupil questionnaire

A simple pupil questionnaire was devised and shared with a sample of intervention schools during the early part of 2020 (between January and March) (see Appendices H and I). The aim of this instrument was to gather further information from Key Stage 4 pupils about their use of FLASH and their attitudes towards it. The questionnaire was designed to be quick and easy to complete for all Key Stage 4 pupils, irrespective of academic ability. The final version of the questionnaire had a Flesch-Kincaid grade level of 4.1 suggesting that it can be accessed by children with a reading age of a typical 9- to 10-year-old. Both an online and paper copy were made available to schools so that they could select which approach was most convenient for them. A total of eight intervention schools agreed to distribute this questionnaire and we received 474 responses. We had originally planned to distribute the questionnaire to case study schools; however, after we approached the schools about this, some declined to respond or explained that they were unable to support with this element of the study. In light of this, the evaluators and development team agreed to contact a small number of non-case study schools (n=5). Two of these schools agreed to distribute the pupil questionnaire to their Year 11 pupils. Schools were asked to distribute the questionnaire to Year 11 pupils. Seven of the schools did this while one school shared the survey with both Year 10 and Year 11 pupils as both year groups were using FLASH Marking. These Year 10 responses have been included within our analysis as they provide important insights from a group of pupils who had been using the intervention since the start of the trial (i.e. the start of their Year 9 programme of study).

It is worth noting here that we began contacting schools about the questionnaire and distributing them to English departments in late-February 2020 and early/mid-March 2020. Due to the Covid-19 pandemic and the lockdown in March 2020, it was not possible to receive responses from all of the schools who had originally committed to sharing the pupil questionnaire.

Observations of training sessions

The evaluation team attended a range of training sessions across the two years of the FLASH Marking intervention. The original protocol does not include plans to observe these sessions; however, it was determined early in the evaluation that this would, be a valuable approach to take for gathering IPE data. This was not included in the updated protocol, which was an oversight on the part of the evaluation team. The first observation was a pilot session during the initial stages of the project, which was conducted with a single secondary school in the North of England in spring 2018. The development team used this as an opportunity to 'test out' the initial training day that they had planned for the intervention schools in July of that year. Four English teachers were present at this training.

The evaluation team also observed a number of sessions from the three main training events for FLASH Marking. In each of the eight hubs, three whole-day sessions were held at different timepoints throughout the trial. The first was held in June–July 2018, prior to the FLASH Marking intervention being used from September 2018; the second was a follow-up session in autumn/winter 2018; and a final session was held at the end of that academic year in July 2019. The number and dates of these are summarised in Table 7.

Table 7: Training sessions and evaluation team attendance

Session and date	Number of hub sessions observed by evaluation team
Training Session 1 (summer 2018)	4/8
Training Session 2 (autumn 2018)	2/8
Training Session 3 (summer 2019)	4/8

Correspondence

During the course of the evaluation, the development team were in frequent contact with English departments using FLASH Marking. They gathered information about the implementation of FLASH via evaluation forms/emails. Where appropriate, and with the permission of those teachers sending the information, these data were shared with the evaluation team to provide further detail and context for the IPE. The information has been analysed thematically—as with the interview data and fieldnotes—and is used to exemplify some of the issues we raise in the process evaluation findings presented below.

IPE analysis

Table 3 (above) provides an overview of the analytical approaches used in relation to each data collection method. Further details are provided here, including information about how the data were triangulated prior to reporting.

Data collected from case study schools (via classroom observations, interviews with staff and pupils, and examination of documentary evidence) and through observations of training sessions was analysed thematically. The data were predominantly collected in the form of field notes in addition to a small number of photographs (for example, of work in exercise books and classroom displays). In line with the logic models, the intended outcomes of FLASH Marking, and the aims of the IPE, the evaluators devised a set of key themes to frame and focus the analysis. Subthemes were then identified, refined, and confirmed by the evaluation team, following a process of coding derived from the rigorous and flexible approach outlined by Braun and Clarke (2006). The coding/thematic framework used can be viewed in Appendix K. Using this framework, brief summary documents were then created for each of the case study schools and training observations. These foregrounded key issues over the period of the trial and allowed us to identify themes, which were recurring across different schools or groups of teachers/pupils.

Analysis of the teacher questionnaire was conducted as part of the wider impact analysis. The lead analyst from the evaluation team used the software programme, SPSS (Statistical Package for the Social Sciences), to support with this, calculating simple frequencies, percentages, and averages to provide an overview of the responses shared by participants in the second survey. The same approach was used for the student survey.

Following analysis of each dataset, data from all sources were collated and reviewed by the evaluation team and examined in line with the logic model/implementation plan, focusing on the key issues of the active ingredients of the intervention, teacher and pupil engagement with relevant activities, and perception of impact. Where more than one dataset was used to answer questions about these elements, a convergence approach to triangulation was used (Creswell and Plano Clark, 2011), integrating the findings from different methods and developing interpretations based upon these. The key findings and conclusions from the IPE are presented below.

Timeline

The trial timeline is outlined in Table 8.

Table 8: FLASH Marking project timeline

Dates	Activity	Staff responsible / leading
September 2017–January 2018	Trial school recruitment	Development team
December 2017–February 2018	Memorandum of Understanding and pupil data collection from trial schools	Development team
February 2018–March 2018	First workload survey completed by trial schools	Evaluation team (supported by development team)
April 2018	Randomisation to intervention/control groups	Evaluation team
April 2018	Schools informed of randomisation outcome	Development team
July 2018	Training Session 1	Development team
October 2018	Case study school selection	Evaluation and Development teams
November 2018–December 2018	Training Session 2	Development team
November 2018–January 2019	Case study visits to schools	Evaluation team
March 2019–July 2019	Case study visits to schools	Evaluation team
July 2019	Training Session 3	Development team
September 2019	Second year of FLASH trial begins	N/A
Ongoing	Support provided to intervention schools with FLASH implementation	Development team
October 2019–December 2019	Case study visits to schools	Evaluation team
February 2020–March 2020	Second teacher workload survey	Evaluation team (supported by the development team)
January 2020–March 2020	Case study visits to schools	Evaluation team
March 2020	Covid-19 national lockdown begins Externally examined GCSEs are cancelled	N/A
23rd March–September 2020	School closures due to Covid-19 for Year 11 pupils specifically	N/A
September 2020–January 2021	Plans to extend FLASH Marking evaluation, communication with participating schools and recruitment to extension	Evaluation team
January 2021	Externally examined GCSEs cancelled for 2021	N/A
February 2021–May 2021	IPE data analysis and reporting	Evaluation team
July 2021	Analysis of teacher workload survey	Evaluation team
September 2021	Completion of draft report and submission to the EEF	Evaluation team

Impact Evaluation

Due to the Covid-19 pandemic, nationwide school closures, and the subsequent cancellation of externally examined GCSEs, no analysis of pupil data for the primary outcome (attainment in GCSE English and English Literature) was conducted. As such, we do not include information here about pupil participants or pupil-level attrition. Details of our analysis for the secondary outcome of the trial (teacher workload) are included below. As background for this analysis, we provide a brief overview of the characteristics of schools included within the trial.

The changes to the trial due to the Covid-19 pandemic mean that the findings presented can only inform our understanding of the intervention on teacher workload. We are not able to draw conclusions about the impact of FLASH Marking on pupils' attainment in GCSE English. While impact on teacher workload is important, and was a clearly specified secondary outcome in the evaluation, it remains proximal to the main primary outcome of pupil attainment (as per the logic model and implementation plan above).

Attrition

Over the course of the two years of the FLASH Marking trial, six intervention schools withdrew from implementing the intervention. One school withdrew in July 2018, following the first training session. Another school withdrew at the start of the academic year in 2018. This was due to having a new Head of Department who did not want to use FLASH. Four further schools then withdrew in September 2019. Again, the start of a new academic year, brought staffing changes, both at department and senior leadership level, resulting in a decision to move away from the FLASH trial. While these schools did not engage with the intervention after their withdrawal, they had consented to their data (and that of their pupils) being used as part of the intention-to-treat primary outcome analysis, and this was confirmed with them following their decision to no longer participate. The schools, however, did not continue with any other evaluation activities and were not sent the second teacher workload survey. No control group schools withdrew from the trial.

Outcomes and analysis

Primary analysis

No primary analysis was carried out due to the cancellation of externally examined GCSEs in 2020 and 2021.

Secondary analysis

This analysis was designed to include the data from the pre- and post-intervention surveys. The first survey received 833 teacher respondents (415 in the intervention group, and 418 in the control schools) from a total of 990 teachers included at the outset of the trial. Responses were received from all 103 schools who had signed up to the evaluation. The outbreak of the Covid-19 pandemic and the ensuing lockdown meant that the post-intervention response was far lower than would normally be expected. Only 358 teachers (159 in the intervention group, and 199 in the control schools) responded to the second survey, and not all of these had completed the first survey.

We ran two analyses. The first, main analysis is based on those teachers providing responses at the outset and after the intervention (218 cases). The second analysis involves all cases with responses to either pre- or post-intervention surveys. However, this is based on two markedly different sets of respondents—the 833 initial cases, and the 358 follow-up cases.[1]

The first analysis is based on 218 matched responses and findings from this analysis are presented below. Respondents in the two surveys were asked for their names, solely in order to match the responses. Only around 50% provided a

[1] The loss of follow-up data will include data lost for the usual reasons, and presumably a larger subset due to Covid-19 and the associated lockdown. The scale of missing responses means that no attempt should be made to adjust the overall results, by weighting or imputation, for example. The loss of data is clearly not random in nature. Therefore, the results of the complete case analysis are presented as internally valid, but not as representative of a larger set of cases. For more on this, see Gorard S. (2020). 'Handling Missing Data in Numeric Analyses'. *International Journal of Social Research Methods*, 23: 6, 651–660.

readable name and, with the low response to the second survey, this means that we were only able to match 218 responses definitively (102 intervention and 116 control). The following tables are based on these cases and provide the fairest (internally valid) comparison between the groups, even though the set of responses is likely to be biased by the low response rates.

Tables 9 and 10 show that the two groups were reasonably well balanced at the outset, in terms of total hours worked and hours spent marking (according to these self-reports). Both groups report fewer hours worked and fewer hours marking after the intervention, but the decrease for both outcomes is clearly greater for the intervention group.

Table 9: Number of hours worked in the last week; matched respondents

	Mean hours total pre-intervention	SD	Mean hours total post-intervention	SD	Gain	SD	'effect' size
Intervention	48.10	12.84	42.98	15.47	-5.13	15.86	-0.16
Control	48.76	12.66	46.15	14.29	-2.61	15.48	–
Overall	48.45	12.72	44.67	14.91	-3.79	15.67	–

SD, standard deviation.

Table 10 shows that the two groups were also reasonably well balanced at the outset, in terms of hours spent marking. Both groups report less time marking after the intervention period, but the drop is greater for the intervention group with an effect size of -0.17.

Table 10: Number of hours worked on pupil marking and feedback in the last week; matched respondents

	Mean hours marking pre-intervention	SD	Mean hours marking post-intervention	SD	Gain	SD	'effect' size
Intervention	8.96	4.69	6.03	4.64	-2.93	5.30	-0.17
Control	9.05	5.57	7.08	5.68	-1.97	6.01	–
Overall	9.01	5.16	6.59	5.23	-2.42	5.70	–

SD, standard deviation.

Data from the second analysis, including all participants who completed either of the two surveys or both of them is reported below. In the pre-intervention survey, teachers in the intervention group reported completing slightly fewer hours of work in total per week than those in the control group (Table 11). This difference, with an 'effect' size of -0.08, may be partly due to different contracts and job status (such as leadership roles), but will be largely sampling variation. The difference is more marked in the post-intervention survey (although these responses are not all from the same cases, see above). For this, the 'effect' size is -0.25. The respondents who were involved with the FLASH Marking intervention reported a noticeably shorter working week (although this could still be due to different contracts and job status).

Table 11: Number of hours worked in the last week; all respondents

	Mean hours total pre-intervention	SD	'effect' size	Mean hours total post-intervention	SD	'effect' size
Intervention	45.11	15.54	-0.08	41.01	17.44	-0.25
Control	46.28	14.44	–	45.32	16.55	–
Overall	45.70	15.00	–	43.40	17.06	–

SD, standard deviation.

For comparison, those staff who only responded to the first survey reported a mean of 44.72 hours (standard deviation [SD] 15.62), with a mean of 45.33 for the intervention group and 44.13 for the control group. This suggests that the dropout group did not differ substantially.

A similar picture appears when considering only the hours worked on marking and pupil feedback (Table 12). The groups were slightly unbalanced at the outset, but the second survey respondents in the intervention group reported substantially less marking than the control ('effect' size -0.27).

Table 12: Number of hours worked on pupil marking and feedback in the last week; all respondents

	Mean hours marking pre-intervention	SD	'effect' size	Mean hours marking post-intervention	SD	'effect' size
Intervention	7.69	5.05	-0.07	5.69	4.30	-0.27
Control	8.03	4.93	–	7.05	5.49	–
Overall	7.86	4.99	–	6.44	5.03	–

SD, standard deviation.

For comparison, those staff who only responded to the first survey (pre-intervention) reported a mean of 7.44 hours (SD 4.81). This again suggests that the dropout group did not differ substantially, although they reported slightly less marking at the outset. This survey provides less than perfect data but taken together, Tables 10 to 12 suggest that FLASH Marking was associated with a decrease in the time teachers spend marking.

Attitudes to workload and FLASH Marking

In total, the second survey was returned by 358 teachers, whose responses are used in Tables 13 and 14. Table 13 shows that the intervention group were somewhat more likely to report that their overall workload was more acceptable than the control group. If we ignore the middle responses, and treat both levels of agreement as the same, then the odds ratio is (19 x 78) / (12 x 61) or 2.02 towards the intervention group.

Table 13: Percentage of respondents in each group agreeing that they have an acceptable workload

Group	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
Intervention (n=159)	1	18	20	38	23
Control (n=199)	0	12	10	47	31

Table 14 shows that the intervention group were somewhat more likely to report satisfaction with the amount of marking they do than the control group. If we ignore the middle responses ('about right'), then the odds ratio is (3 x 60) / (3 x 37) or 1.62 towards the intervention group.

Table 14: Percentage of respondents in each group making each statement about the amount of marking they have to do

Group	Too little	About right	Too much	Far too much
Intervention (n=159)	3	61	27	10
Control (n=199)	3	37	38	22

In the second survey, 159 teachers responded from the intervention group. They tended to agree that the training they received to use FLASH Marking was helpful, it had reduced their marking workload, and that they would encourage other schools to use it (Table 15). Perhaps most importantly, 77% claimed that the intervention had benefited the pupils.

Table 15: Percentage of respondents (n=159) in the intervention group agreeing with each statement

Statement	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
The FLASH training was useful	51	26	12	7	4
FLASH has reduced their marking and feedback workload	44	32	10	11	4
They would recommend FLASH to others	55	24	9	8	4
Pupils in Year 11 have benefited from FLASH Marking	46	31	10	8	5

In summary, the FLASH Marking teachers reported useful training, are somewhat happier with their workload, report a lower workload including marking, feel that the impact for pupils was beneficial, and would suggest to others that they use this approach.

Implementation and process evaluation results

In this section, we discuss the findings from the IPE. In more usual circumstances, we understand that these would be used to exemplify and explain the findings from the impact analyses. However, without the primary outcome data (from pupils' outcomes), this is of course not possible. Despite this, we still feel that there is considerable value in reporting key findings from the IPE. First, they provide a number of helpful insights, which may be useful for the funders and other development/evaluation teams to be aware of in future trials in the area of feedback and marking, and which take place over an extended period of time. Second, we are able to link some of the IPE findings to our secondary outcome data surrounding teacher workload. Again, this may be helpful for informing future research in this complex area. Finally, at the outset of the FLASH Marking project, the evaluation team set out with a view to conduct an in-depth and rigorous IPE, which sought to understand the intervention and its implementation in some detail. We designed a process evaluation, which ensured that teachers' and pupils' voices could be heard. Moreover, our IPE included a longitudinal element, including multiple visits to case study schools and regular contact with some of the staff involved. As a result of this, our IPE dataset offers some quite unique and important perspectives about using an intervention such as FLASH and being involved in an evaluation trial for a period of two academic years.

Our IPE findings are organised as follows. First, we discuss the implementation of FLASH Marking, beginning with an outline of the compliance data that was collected. As part of the original evaluation protocol, we had planned for this to be correlated with attainment information. However, this is not possible due to the lack of primary outcome analysis for this study. Instead, we describe the compliance findings here and draw on them to inform our discussions of implementation and process. Next, we discuss in more detail the implementation of the intervention, including fidelity to the intervention and factors, which appeared to influence the implementation in schools. Second, we include details of findings relating to the *perceived outcomes* of FLASH Marking on both the primary and secondary outcomes (pupil attainment and teacher workload) as well as on other outcomes described by those involved with the programme. For both of these sections, we use the logic model and the implementation plan (Figures 4 and 5) as frameworks to inform the IPE of FLASH in schools. Finally, we discuss the potential *sustainability and scalability* of FLASH Marking.

Implementation of FLASH Marking

Compliance

As outlined in the SAP, compliance data was collected based on three key areas:

1. intervention schools' attendance at all three training sessions;
2. confirmation that FLASH training was cascaded to English teachers prior to the start of the trial in September 2018 (following Training Session 1); and
3. the extent to which Heads of Department felt that their department had committed to FLASH implementation in the first 15 months of the trial.

The first of these issues was measured by registers collected at training sessions by the delivery team. A summary of attendance at each of the training sessions is included in Table 16. This shows the total number of attendees at each training session, how many participating schools were represented at the training, and the number of schools who had withdrawn from the implementation or that were only able to send one teacher to the session.

Table 16: Attendance at each of the FLASH Marking training sessions

Training session	No. of attendees	No. of schools represented	No. of schools sending just one teacher	No. of schools withdrawn and sending zero teachers
Training Session 1	103	52	1	0
Training Session 2	98	49	0	3
Training Session 3	92	46	0	3

The second set of data was collected via email confirmations from trial schools to the development team. They received confirmation that all trial schools (n=52) delivered cascade training to all Year 10 English teachers in intervention school English departments. In four cases, the development team were asked to visit the schools to support with this.

Finally, English Heads of Departments were asked about their commitment to FLASH in the second teacher workload survey. Of the 52 intervention schools who were sent the second workload survey, 40 Heads of Department responded. The 12 schools with no Head of Department response included the six schools who had withdrawn participation during the first year of the trial plus a further six schools.

An overview of the Head of Department responses to the compliance question is provided in Table 17.

Table 17: Response to item 'My department have been fully committed to FLASH Marking since the start of the trial'

	Strongly agree	Tend to agree	Neither agree nor disagree	Tend to disagree	Strongly disagree
Number of schools (of 40) (%)	11 (27.5%)	25 (62.5%)	3 (7.5%)	1 (2.5%)	0 (0%)

These data show that the majority of Heads of English reported satisfactory levels of department commitment to FLASH Marking for the first 15–16 months of the trial. Of course, it is possible that these data are somewhat skewed by more committed Heads of Department/departments completing the survey and those less committed choosing not to respond. Nevertheless, taken as a whole, the three compliance measures indicate generally positive engagement with FLASH Marking, particularly in the earlier stages of the trial. This corresponds with additional fidelity and implementation data that we gathered from participating schools and the development team.

Fidelity

As can be seen in the implementation plan (Figure 5), the fidelity of FLASH Marking was conceived via four core 'active ingredients'. These were viewed as key elements of the intervention, which in theory and if delivered as intended, would contribute to the specified outcomes of the trial of improved pupil attainment and reduced workload for staff. The four ingredients can be summarised as:

- removal of number/letter-based grades from day-to-day feedback;
- using FLASH codes as part of English lessons, for example, in success criteria, learning outcomes, and model examples;
- FLASH codes are used by teachers to provide personalised feedback to pupils and to inform planning of lessons; and
- FLASH codes are used to support pupils' metacognition, for example, through reflection and improvement tasks, and target-setting for future work.

Through our visits to case study schools and incorporating information from teacher and pupil surveys and correspondence with the development team, we found that these four ingredients tended to play a central role in the implementation of FLASH Marking. These elements were shared explicitly with English department staff at each of the training sessions and reiterated by the development team as part of their follow-up support work too. As a result of this focus on the core ingredients throughout the course of the trial, we have used these to underpin our analysis and reporting in relation to the fidelity of the trial. This is a slight deviation from the protocol and the SAP where compliance and fidelity are discussed more interchangeably. We acknowledge, however, that it is vital to examine these fidelity issues separately, and in some detail, and thus outline our core findings below.

Removal of number/letter-based grades

In all of the case study schools that we visited, teachers reported that the grading of pupils' formative work (for example, written tasks completed in exercise books or practice exam questions completed in class or for homework) had been reduced for the target cohort. Most schools still reported retaining some element of grading though, usually for more formal assessments such as mock exams. Examination of exercise books and discussions with pupils corroborated this view, particularly during the first year of the trial.

During the second year, some English departments did appear to be using number/letter-based grades more frequently as highlighted by responses to the pupil survey. When asked if they are given marks/grades in addition to other forms of feedback, 50% of pupils (237/474) reported that they 'Always' or 'Often' receive a numerical mark/grade. A third (151

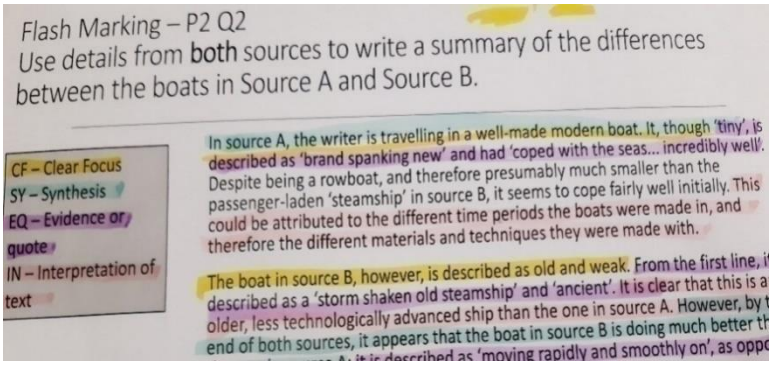
pupils) said that they were 'Sometimes' given a grade, with just 5% (24 pupils) saying this 'Never' happened. Interviews with teachers indicated that they were sometimes required by school policy to provide regular grades/marks to pupils, especially those preparing for their GCSEs. While they were generally pleased to have shifted the focus to more formative feedback approaches, most acknowledged that removing grades altogether was sometimes not easy or feasible in their current contexts. One Head of Department reflected this wider view noting that it had been 'a bit of a challenge to wean our team off marking' (referring to detailed comments and grade-focused marking), but stated that the FLASH principles became more embedded within the first two terms of the trial 'as the benefits in terms of reduced workload and students' progress are becoming apparent'.

Using the codes in the classroom

Observations and interviews across case study schools and resource examples collected from other intervention schools indicated a widespread use of the FLASH Marking codes in classroom-based activities. Visits to schools highlighted the varied and innovative ways, which teachers were integrating the codes into their lessons. Teachers were given considerable freedom to incorporate FLASH codes into their teaching in any way that they deemed appropriate. As we note above, the development team provided some examples of how this could be done but English departments were also encouraged to think and act creatively in their use of the codes.

We saw examples of codes being used by teachers to highlight lesson objectives, to provide success criteria, or as feedback to pupils. We also saw pupils frequently using them, usually for formative assessment purposes. The images below in Figure 6 show examples of resources created by teachers in intervention schools.

Figure 6: Images of example resources using FLASH Marking codes



(Pupil work, Dickens Hub, London)

(PowerPoint, Russell Hub, North-West)

In Figure 6, the image on the left shows FLASH codes being used as part of a 'model' writing exercise. This was something that we saw regularly across the case study schools, with teachers asking pupils to identify where they could find evidence of each of the skills. The image on the right shows an example of FLASH codes being used within a PowerPoint slide. Here, the teacher was using the codes to highlight the focus for the lesson and to reiterate the skills and knowledge that would be focused on that day. During our observation of this lesson, the teacher explained what the codes meant and how they aligned with the topics being covered (in this case, poetry from the GCSE English Literature course). She also used the codes on this slide and careful questioning to check pupils' understanding of the kinds of skills that they would need to demonstrate in their upcoming exams.

As we can see in the first image on the left in Figure 6, tasks using FLASH codes were often accompanied with colour-coding to support pupils to identify and delineate between the different elements or skills demonstrated in a written text. As an alternative task, pupils were also frequently asked to complete a piece of writing and then use the colour-coding/code identification to self- or peer-assess, which of the skills they had included or accomplished within the work. Activities such as this had been shared and promoted as part of the training and support with the development team. They were designed to facilitate consistent and accurate code use, and to encourage pupils to reflect on their work and to set targets for improvement too. One Head of Department commented on how this kind of regular and embedded approach to using the codes had led to improvements in pupils' understanding of their progress.

> *Our students across key stages now take ownership of their own work and improvement processes. We are increasingly finding students are not only able to identify their own strengths and weaknesses but*

are able to accurately band/mark their own work against GCSE mark schemes. This was beyond our dreams previously.

(Head of Department, Shelley Hub, South-East)

The colour-coding emerged as a particularly positive element of FLASH Marking, especially from the perspectives of the pupils involved. They tended to speak enthusiastically about the use of highlighters to identify the different features of their writing and to spot 'gaps' or areas for development. Staff also felt that there was value in the colour-coding for the reasons described above. However, practically speaking, some noted that it was occasionally a distraction for some pupils. In addition, it was not always possible to ensure a constant supply of highlighter pens for pupils to use, usually due to departments working with very tight financial budgets.

Pupils that we spoke to tended to like the code sheets that they had in their exercise books, which acted as a reminder of what each code stood for. Referring to these code sheets was helpful for learning the codes and for applying them to relevant activities. Each school was required to have a code sheet, and to ensure it was shared with pupils. This often included them being stuck in exercise books and used as displays within classrooms.

Figure 7: Adapted code sheet excerpt, Shakespeare Hub, Midlands

Some schools adapted the original sheet devised by the development team, as with the example in Figure 7 above. Generally, this was due to a feeling that there were too many codes overall and so they needed reducing, or that the original codes did not align clearly enough with the exam specification that the school was working with.

The number of codes is too high and can cause confusion. We have reduced the number of codes used and agreed on a core set of codes by elements of schemes of work…Overall, I think the system is great but feel that the codes need to be streamlined.

(Teacher, Coleridge Hub, South-West)

There was some variation in the frequency and quantity of FLASH Marking code use both within and across schools that we observed. In some of our case study schools, FLASH was apparent in every lesson that we saw and appeared regularly in pupils' exercise books. In other departments FLASH was less visible. A teacher at one case study school in the Shakespeare Hub noted that while he was very enthusiastic about FLASH and could see its potential, other, more experienced teachers in his department were using it much less and had been reluctant to embed it consistently into their regular teaching practice. In this school and one other that we visited, some teachers explained that they did not always use FLASH Marking with 'lower ability' pupils or those with Special Educational Needs (SEN). They suggested that the number of codes plus the need for the pupils to understand a new system of marking and feedback would be too challenging. In the majority of our other case study schools, however, we observed examples of FLASH being used with lower attainers in English, and a number of teachers were very enthusiastic about the opportunities and outcomes that they saw for these pupils.

No doubt about it, FLASH has saved my bottom set Y11 [Year 11] class. They are a bright bunch but have always lacked self-confidence. Using FLASH as a structure is allowing them to feel secure in knowing how they should respond to questions with a structure that isn't too prescriptive. They are now a lot better at reviewing their own work independently and can pick up on the positives as well as what needs improving. Their greatest

> *achievement with FLASH is the fact that they now don't just pick up the WHAT but the HOW, and this is allowing them to redraft work in a much more meaningful way. I have just marked their mocks and they have all used the FLASH structures which I have taught and it is their best bit of work to date.*
>
> (FLASH Marking Lead Teacher, Wordsworth Hub, North-West)

Teachers at a school in the Dickens Hub, London, shared their view that FLASH Marking 'has been more beneficial for low-ability groups as students can improve their vocabulary and comprehension'. However, another colleague from this same school commented that she had not introduced FLASH Marking with one small class of boys who had a number of behavioural issues. She felt that the approach would be too difficult to explain and embed with a group such as this.

Curriculum and planning

As noted in the logic model and implementation plan, there was an expectation that FLASH Marking should become embedded within the medium-term and longer term planning of each school's Key Stage 4 English curriculum. A number of the strategies discussed in the sections above (coaching, visits to schools, and resource-sharing, etc.) were designed to support with this, and to integrate FLASH into the subject for both teachers and pupils. Our observations and discussions with English staff indicated that in some of our case study schools, FLASH was being added into schemes of work (i.e. medium-term/longer term curriculum plans) to highlight the skills and knowledge that teachers should be focusing on at each point through the two-year GCSE course. While this did not seem to be the norm across all schools, a number of teachers during the school visits did indicate that they could see value in this kind of longer term planning. In two case study schools during a visit in Year 2 of the trial, staff commented that they would like to embed FLASH more fully into their scheme of work documents but that this had not been possible at that point due to time constraints.

In some schools, rather than focusing on schemes of work, teaching staff had adapted their lesson resources (such as PowerPoints and worksheets) to include reference to FLASH Marking codes. Many schools were using a common set of resources for their delivery of units of work, meaning that all teachers in the department would use the same materials (perhaps with adaptations) and the majority of children in the cohort would see the codes being used in their lessons. This approach appeared to be a more concrete, student-facing strategy, which facilitated the sharing of the codes and their use within specific lessons.

Supporting pupils' metacognition

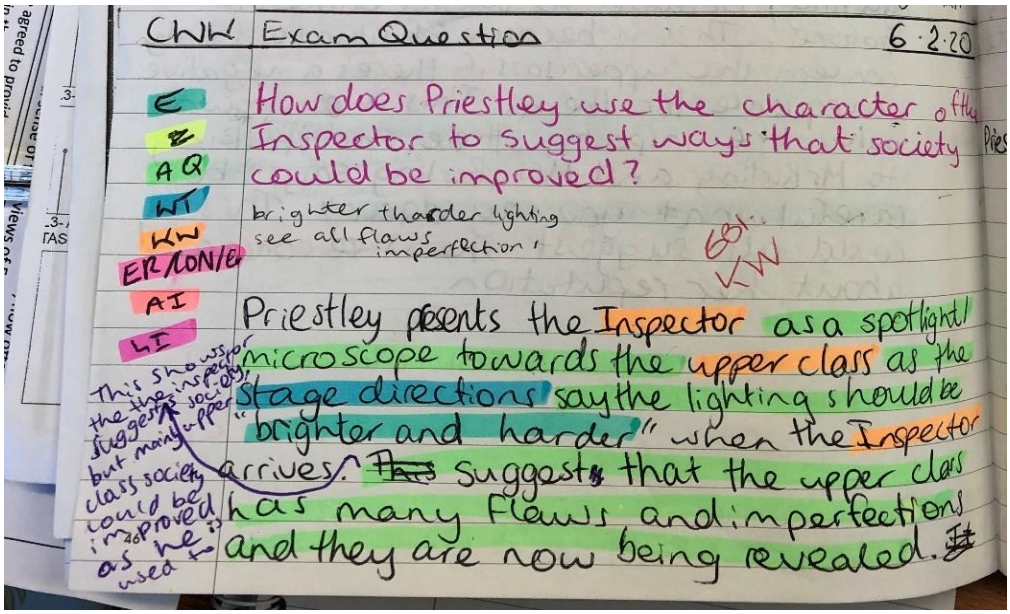
The fourth 'active ingredient' refers to improving pupils' metacognitive skills. Evidence of this element was noted in relation to some of the sections above (for example, using the codes in the classroom, and curriculum/planning approaches). Our IPE data also indicate that the FLASH Marking intervention explicitly supports the development of these skills by encouraging pupils to think independently about their learning and progress through the processes of planning, reviewing, and improving their work. During the training sessions, the development team made frequent reference to metacognition, sharing contemporary research and evidence on the topic (such as the EEF guidance report), and supporting teachers to consider how this could be applied in the classroom. In the third training session, particular emphasis was placed on metacognitive skills. Activities involved asking teachers to reflect on how they had been using FLASH Marking to promote metacognition, how they were teaching these skills to pupils, and the potential for the intervention to support metacognitive talk in the classroom. Short video excerpts were used from a small number of participating schools to highlight some of the different strategies used. Teachers across the hubs that we visited responded very positively to this session and reported that FLASH either aligned with their approaches on metacognition more generally, or that it had supported them to promote this focus with colleagues and pupils.

In the classroom, we saw some evidence of the approaches noted in the implementation plan relating to metacognition. As we note above, the highlighting/identification of key elements within pupils' writing was one of the most common strategies that we saw. During our first round of school visits in autumn/winter 2018, we saw evidence of this approach in 12 out of the 13 intervention schools that we attended. Most of the lessons that we observed in this round of visits included some element of highlighting. In the three schools where we did not see this, there was evidence from pupils' exercise books that they had used this approach in recent weeks. In subsequent visits, we saw similar, frequent use of this highlighting technique. A small number of teachers/pupils moved away from multi-coloured highlighting later in the trial, instead sometimes using a single-coloured pen for identifying features or relevant aspects of the text. The principle was still similar though in that pupils were being expected to reflect upon the content of their writing, identify features correctly, and address where there were gaps or errors in their work. From our observations and discussions with

teachers, we found that the skill of highlighting/identification needed to be developed carefully with pupils, before they were confident and able to highlight the correct elements aligning with relevant FLASH codes. Over the course of the IPE visits, we saw examples of teachers providing guidance on this for pupils, often using model examples to indicate the kinds of content that pupils should be including in their writing and identifying in their work or that of others.

The extract from a pupil's exercise book, in Figure 8, provides an example of how the codes were being employed as part of the planning, feedback, review, and improvement process. In this example, the focus codes were shared by the teacher and the pupil wrote them in the margin as a reminder while completing their practice exam question. The codes were then colour-coded and the pupil highlighted relevant sections. The next step involved the teacher taking a look at the work and providing an EBI (Even Better If) target, also using a FLASH code. Finally, the pupil returned to the work and sought to improve and extend their response (as can be seen in Figure 8 by the additional sentences included in the margin). This approach focused on written tasks as a 'work in progress', encouraging pupils to constantly aim for development and improvement.

Figure 8: Example from a pupil exercise book, Bronte Hub, East Midlands



Examples such as the one above in Figure 8 were common in the case study schools where we carried out observations. By the time of our visits in the second year of the trial (2019–2020), there was a sense that this was a well-embedded approach for many teachers and pupils in FLASH schools. A number of the teachers we spoke to noted the power of this strategy for supporting pupils' independence and for promoting a focus on continual improvement. In some schools, the development of metacognition skills was closely linked to the use of self- and peer-assessment. The former appeared to be a more popular approach with some teachers suggesting that it was more time efficient and that pupils preferred to think about improving their own work rather than that of their peers. In one school, however, a teacher shared that she embedded peer-assessment through an alternative approach, using pupil-written examples as models to be shared with the whole class for evaluation and discussion in line with the FLASH codes.

Factors influencing implementation

In addition to the partial list of 'active' ingredients described above, our analyses also examined other factors, which appeared to influence the implementation of FLASH Marking in schools, and so should be added to that list. A number of elements relating to different stages of the intervention, including staff training, support from the development team, staff engagement and availability, support from school leadership, and teachers' perceived workload were found to influence implementation (either positively or negatively). We describe our findings on each of these factors below, drawing upon evidence from the training session observations, the case studies, the staff questionnaire, and correspondence shared by the delivery team.

Training for staff

The delivery of FLASH Marking was underpinned by a set of three day-long training sessions with English department representatives from each intervention school. Observations of these sessions across different locations and with different hubs of schools, plus feedback from information shared by the development team, indicates that the professional development element of FLASH Marking was very well received. Teachers involved, reported that the training was high-quality, well-prepared, and informative. Many really valued the way that the content was underpinned by research evidence and practice in order to explain and justify the approaches taken during the development of FLASH. The observed training sessions incorporated a mix of trainee-led (i.e. by the development team) content linked to the principles of the intervention and the wider issues associated with these, such as assessment, metacognition, and promoting pupil independence. In addition, opportunities were built in for teachers to share examples of practice from their school and to consider how the training could be applied in their contexts. This allowed for discussions of good practice across the hub groups, and was received enthusiastically, particularly in the earlier stages of the trial when teachers were still considering how best to embed FLASH Marking into their practice.

As we note above, attendance at the training was very good throughout. Teachers were pleased that they could attend Continuing Professional Development (CPD) out of school for a day and that the costs were covered by the incentive funding that their school received for participating in the trial. Participants were consistently very positive about the professional delivery of the training and the friendly and knowledgeable approach taken by the development team. The evaluations and feedback following each session were also found to be useful for informing support for schools and future training too.

Following the initial training session in summer 2018, those teachers who had attended were expected to cascade the information they had received to the rest of the English department. We note above that all participating schools reported doing this, although discussions with staff on school visits suggested that there was some variation in the delivery of this cascading, the prioritisation given to it, and the extent to which teachers felt ready to implement it effectively in the first term. Two Heads of Department noted the challenges, which arose when new members of staff joined the school as they felt obliged to re-deliver the cascade training, but did not always have the time to do this in as much detail as they would have liked. In another school, we noted that the teaching assistants (TAs) were not familiar with FLASH but were supporting pupils who were using the codes. The Head of Department at this school was hoping to cascade some elements of the FLASH training to the TA team to support with their work with young people in English lessons.

Support, monitoring, and resources from the development team

In addition to the more formal training sessions described above, the development team also offered regular support to intervention schools, including visits to schools. This encompassed a number of different elements including visits to English departments, regular communication with schools, and resource-sharing via the online Trello platform. Engagement with these approaches was primarily left to schools to determine their needs and the level of support required. Visits to school were carried out as part of the fidelity-checking approach from the development team, and for providing additional support and guidance to schools who needed it. School visits were popular with many schools, with 32/52 intervention schools welcoming the development team at least once during the trial (until March 2020). A further five schools had visits scheduled, which had to be cancelled due to the Covid-19 pandemic.

As part of their visits, the delivery lead for school visits, typically engaged with some monitoring activities (for example, book scrutinies or lesson observations) and some coaching/mentoring for individuals or groups of colleagues. The comment below from a Head of Department exemplifies the value that teaching staff reported in relation to these visits:

> [*The FLASH development team] visited us for monitoring and support. They spent the day in the English department, observing lessons, talking to students and teachers and conducting book scrutinies. They also stayed for our department meeting to feedback to us all and discuss how we felt the trial was going…The feedback that we received was excellent, and to get external verification of our implementation of the trial was really reassuring as this process was obviously new to us.*
>
> (Head of Department, Wordsworth Hub, North-West)

The comment above, is indicative of the positive response to the support received by the development team. English teachers tended to report increased feelings of confidence with FLASH Marking and a commitment to ensuring that the intervention was being delivered as required. Feedback provided by the development team was viewed as constructive and practical, encouraging departments to amend or develop their practice where needed. Arguably, this led Heads of

Department and teachers to take further ownership and responsibility for their delivery of FLASH, knowing that they were able to utilise the support from the development team if needed.

Schools were also asked to remotely share examples of work using FLASH approaches with the development team. These provided insights into the different ways that departments were using FLASH Marking (via lesson plans, schemes of work, resources, pupil work, and teacher feedback), and allowed the developers to observe where schools might be having difficulties or where they may have misinterpretations of the training or the intervention. Again, schools were positive about this. After the second training session, Heads of Department were asked to send a document, which indicated how consistency of FLASH was being monitored in their school. Of the 49 schools using FLASH Marking at this point, 48 completed this task and received feedback from the developers on their contributions. Following the third training session, all schools were asked by the development team to send an example of a document demonstrating how pupils were using FLASH (for example, peer-assessment tasks and exam responses). Of the 46 schools implementing FLASH at this stage, 42 sent these documents. Some schools shared examples of good practice with FLASH more frequently, often via the online Trello platform and through the development team's 'FLASH Friday' initiative, which involved highlighting positive or innovative examples of FLASH use to other schools in the trial once per week. We did not collect information from the Trello site about interaction or engagement with these shared resources. At Training Session 2, some attendees reported that they found the Trello examples interesting or helpful; teachers were more positive, however, about the sharing of resources via the weekly FLASH Friday emails. One teacher, for example, stated that she liked the regular updates and ideas, which appeared in her inbox, and that she frequently adopted these in her ongoing teaching practice.

Despite some engagement with the Trello platform, overall, our data suggest that teachers' views of it were varied; our discussions with school colleagues suggested that many did not view it as a core part of the FLASH intervention. Our observations from training sessions indicated that while the development team encouraged the use of Trello, it was presented as more of an optional part of the intervention. This was reflected in teachers' views of it. Some teachers explained that much of the content on Trello related to one exam board; if they were working with another board or specification, they felt that the resources were less relevant and so were less likely to engage. At one of the training sessions, two teaching colleagues also reported that Trello was yet 'another platform' that they needed to learn how to use, and that its potential benefits did not outweigh the time needed to engage with it in the intended way. As a method for monitoring the delivery of FLASH Marking, however, the development team found the online platform useful. It provided an easy way to assess the kinds of resources and activities that intervention schoolteachers were using and enabled timely feedback and advice where needed.

Throughout the trial the development team frequently communicated with participating departments, usually via email. This correspondence was used to reiterate the support available to schools, to share ideas, updates, or resources, and to promote a more general sense of motivation and engagement with the intervention and the trial. The development team also reported that this was an effective approach for maintaining contact with the majority of schools, and keeping contacts current, especially where staff turnover was an issue.

Staff availability and engagement

At the outset of the FLASH Marking project, staff engagement with the intervention was high. This was important as the development team were relying upon whole English departments to take-up FLASH Marking and to embed it with fidelity into their day-to-day teaching. Without this whole-department approach, any intervention impact (positive or negative) could be affected. We note above, the strong compliance via attendance at the training sessions, and the willingness of schools to designate two English teaching colleagues (including the Head of Department) as FLASH Marking Leads. Observations at the training sessions confirmed the enthusiasm for and commitment to understanding and implementing FLASH Marking. Teachers were eager to learn about a new strategy, which was underpinned by research and evidence, and which might positively influence academic outcomes and workload. During the first year of the trial, our visits to schools indicated that, for the most part, English departments were continuing to engage with FLASH and to use it within lessons. In some schools, the opportunity to adapt FLASH Marking to create resources and to develop schemes of work using it, were very much welcomed and led to many colleagues embracing the intervention and promoting its use within the department (and sometimes beyond). In some schools, it took longer to encourage all teachers to embrace FLASH Marking and to change their traditional marking habits, as this Head of Department commented in the second teacher survey:

> *Getting teachers to use a new way of marking has been the biggest challenge—there is an engrained dependence on written comments/feedback, and FLASH also requires a quick reading speed to really reduce workload, I think. However, those who have embraced it love it! Those who have used it a bit less regularly and maintained some older habits alongside have found it more challenging to implement—because it's not in regular use. When it's being used consistently it's worked extremely well.*
>
> (Head of Department, Burgess Hub, North-West)

Some recognised it as a way to showcase their teaching and learning skills and innovation. One FLASH Marking Lead teacher, for example, from a school in the Shakespeare Hub, described how he had used the introduction of FLASH and his leadership of its implementation to support a successful promotion application for a whole-school role. This additional motivation to 'make FLASH work' could have been important for encouraging others to use the intervention as intended over the course of the trial.

Engagement on both an individual teacher and wider department level did appear to be affected by staff turnover. With a two-year trial, it was expected that a number of teaching colleagues from across the project would change roles or leave the school where they were working at the start of the project. Our findings from the workload survey (see above) highlight the fact that many teachers did indeed move on, making way for new colleagues to join the school part-way through the FLASH Marking trial. Two key issues arose from this. First, delivery of and engagement with the intervention were sometimes lost where colleagues who were involved in the leadership of FLASH left the school. This became more problematic if additional teachers from the original FLASH-trained group left. Second, when new English teachers arrived, there was not always capacity or willingness to train them in the intervention or to support them to implement it with their classes. In schools where there were, for example, department-wide schemes of work and resources this was less of an issue, but where teachers were working more independently on their planning and teaching, FLASH could go unnoticed and unused.

Six of our case study schools experienced a change in leadership within the English department. Our observations and discussions, with both teachers and the development team, pointed to a decline in the use of and engagement with FLASH Marking as a result of this. New Heads of Department tended to join the school with their own agenda for improving attainment and were sometimes reluctant to learn about a new approach, which they were unfamiliar with and had not introduced themselves. In some of the schools, existing teachers tried to continue using the intervention where they could, but their comments suggest that they were not able to maintain the level of implementation that they had previously had.

The development team worked very hard to offer support and guidance where there were staffing changes in FLASH Marking schools. They offered and attended visits, delivered refresher sessions to new colleagues, where required, and communicated frequently with new English subject leaders. This worked well where schools (and English departments) remained committed to the trial; however, in some cases staff changes led to a lack of contact with the development and evaluation teams, and presumably limited engagement with the intervention too. This is likely to have affected overall implementation and perhaps outcomes associated with the intervention too.

Support and engagement from senior leaders

At the outset of the FLASH Marking trial, each participating school was required to nominate a representative from the senior leadership team who would act as a further contact during the project, and would support the implementation of the intervention with English departments. We found some variation in the levels of involvement of these senior colleagues and the amount of support that they provided to the English team. Where senior leaders were also members of the English department, we tended to see increased 'buy-in' with FLASH Marking. In some schools, teachers described being 'left alone' by senior leaders to roll-out and implement FLASH Marking whereas other case study teachers reported that their leadership team were more involved and supportive. A number of case study schools explained that senior leaders were involved in assemblies about FLASH or spoke regularly to teachers and pupils about the intervention. Where senior leadership were significantly involved with supporting FLASH Marking, English Heads of Departments and teachers spoke positively about this and felt that it made a difference in terms of instilling expectations around the intervention and ensuring that pupils and parents were clear about the project and its aims. While this engagement was welcomed by those English departments who had received support from senior leadership, not all

senior teams were as aware or engaged with the FLASH trial. Heads of Departments, however, did not report this as a particular challenge. Evidence from the interviews suggests that they viewed input from departmental staff (i.e. English teachers and support staff) as more valuable than that of those from beyond the English department.

As we see in the section above, where leadership personnel changed, the emphasis on and engagement with FLASH Marking sometimes changed too. In two case study schools, for example, a new head teacher with different school improvement priorities led to the schools withdrawing from FLASH implementation in the second year of the trial. Sometimes these decisions were reportedly linked to additional factors such as GCSE results or Ofsted outcomes, but they tended to be met with disappointment from English teachers who had invested time and effort into the study.

Time and staff workload

Issues relating to time and workload were integral to the FLASH Marking intervention and implementation. With workload as a designated outcome of the trial, staff were very aware that this was an approach designed to reduce the time spent on marking young people's work, and this was often an issue that they were keen to discuss with the evaluation and development teams.

Time for understanding the intervention and planning for its implementation was a core element of the training days led by the development team. During these sessions, teachers were invited to consider when and how they would cascade the training to their colleagues, and what time would be needed for monitoring, additional training, and support. The development team used their training materials to outline the time commitment needed to ensure that FLASH was embedded in a meaningful way.

Back in school, it was generally clear that during the early stages of the trial, time was allocated for cascading the training, supporting colleagues with its implementation, and monitoring use. Heads of Department also saw value in coaching and support visits from the development team, often requesting more than one of these as a way of supporting and promoting professional development. In one school, a Head of Department also described spending a lot of time in the summer holidays, before the trial started, reorganising and amending Key Stage 4 schemes of work to include the FLASH codes, and to ensure that they were ready for a new school year with three newly qualified teachers joining the department. Discussions with teachers though did point to general workload issues sometimes impacting on their ability to employ FLASH in the way that it was intended. A small number of teachers referred to the amount of time it took to introduce FLASH Marking to pupils. They were concerned that time spent on this meant that they were not able to spend time on other planning, teaching, and assessment duties. However, on later visits to schools, most teachers acknowledged that while time was needed at the outset to embed FLASH, once that had been done (by staff and pupils) then time could be used for other activities. One teacher remarked that one of the key benefits of FLASH Marking was that because the pupils were self-/peer-assessing their own work more and finding the 'big things' to focus on, teachers could spend more time examining more specific issues or areas for development.

Perceived outcomes

<u>Pupil outcomes</u>

Both teachers and pupils were generally positive about the potential impact that FLASH could have on academic attainment. The second teacher survey, carried out during Year 2 of the trial, indicated that 77% of respondent teachers (n=122/159) felt that FLASH Marking had had some positive benefits for pupils. Visits to schools and discussions/correspondence with teachers tended to support this view and provided further insights into the kinds of benefits that teachers identified. Perhaps unsurprisingly, those teachers who were enthusiastic about using FLASH and who appeared to have embedded it well within their lessons, seemed to believe that it was also having a positive impact on pupils' learning and attainment. One Head of Department from a case study school in Shelley Hub, London, explained how the FLASH codes that their school used helped pupils to recognise what skills and knowledge were required in the exams; this, he said, led to them understanding what they needed to include to gain marks and to improve on previous performance. In a non-case study school, this English subject leader reported, via the survey, that pupils' outcomes were improving as a result of the impact of FLASH on teaching and learning:

> *Flash Marking is improving teaching and learning because staff are more likely to adapt and plan lessons based on in-class assessment and so lessons are more tailored to students' needs.*

(Head of Department, Burgess Hub, Manchester)

In correspondence between another senior leader/English teacher and the development team, this teacher felt that FLASH was leading to improved outcomes in relation to assessments, supported with positive views from pupils too:

> *My Y11s [Year 11s] (lower ability) have just completed their mock exam and have done so well—they were using their FM [FLASH Marking] codes to plan and structure their work and have all commented on how helpful it has been.*

(Teacher and Senior Leader, Shelley Hub, London)

These perceptions of improved attainment were not uniformly held by all case study school English teachers. In some schools, we spoke to teachers who were unsure of the academic benefits of the intervention. These comments, however, were often associated with concerns that either pupils or the teachers themselves were not implementing FLASH Marking properly. Sometimes these teachers did not feel confident relying upon FLASH Marking and so combined it with other approaches (for example, more written feedback or marks), as exemplified in this survey comment:

> *The impact on pupils' learning has been minimal, as most pupils fail to engage with the programme effectively, and see the need to flash mark their work more of a chore than a useful tool. Many pupils fail to see how the codes effectively correspond to their work. The impact on marking work is minimal—it has been poorly communicated how marking should look when used effectively with the codes, and I tend to mix old-style written feedback with codes which, if anything, adds to my workload.*

(Teacher, Burgess Hub, North-West)

As described, this combination approach clearly has potential implications for pupil attainment and teacher workload. While this view was in the minority (according to both our case study and survey data), this teacher does highlight the need for staff to feel confident with the intervention and its potential impact if they are to implement it as planned.

In addition to teachers' views on pupil attainment, we also collated pupil perspectives. Through the pupil survey, young people reported generally favourable views about the intervention and its influence on their learning (see Table 18).

Table 18: Pupil views on FLASH Marking and impact on learning (n=474 pupils)

	Strongly agree (%)	Agree (%)	Neither agree nor disagree (%)	Disagree (%)	Strongly disagree (%)
FLASH Marking has helped me with my English work	19	45	23	10	4
FLASH Marking helps me to know what to improve in English	18	55	16	6	4

In our case study schools, pupils that we spoke to expressed a range of views. Most understood the purpose of FLASH Marking and often talked about the value of using feedback to improve their work. In line with some of the teacher perspectives above, and the pupil survey, many pupils felt that FLASH offered a clear, focused approach to improvement and development. This, they felt, was often supported by tools such as the colour-coding/highlighting and the use of model responses. The comments below exemplify some of the positive views from Year 10 pupils in our case study schools during the first year of the trial:

> *[FLASH Marking] gives the chance for teachers to explain what I need to work on in an assignment as the explanations are centralised*.

(Year 10 Pupil, Shelley Hub, South-East)

> *The highlighting clearly displays what we've done well or what we haven't included.*

(Year 10 Pupil, Shelley Hub, South-East)

> *[FLASH] is a way to be aware of what you are writing, and said that it means you don't miss anything out.*
>
> (Year 10 Pupil, Russell Hub, North-West)

A small number of pupils reported that they did not find FLASH Marking helpful for their progress in English. This view was nearly always expressed when comparing FLASH with other approaches. These pupils, who were typically found in higher attaining groups, explained that they preferred to receive longer comments and/or marks/grades too. They tended to find FLASH too 'neutral' in its approach and wanted more positive affirmation of the work they had done and the achievements they had demonstrated. A higher attaining pupil at one of our case study schools made the following comment:

> *It is quite overwhelming, we get told to do things in a certain order, which makes it harder to be creative in my writing as you have to follow the guidelines.*
>
> (Year 10 Pupil, Dickens Hub, South-East)

This was not a view that we heard from many other young people; however, concerns around FLASH restricting creativity with writing were also voiced by two teachers from two different case study schools. It is worth noting though that these were comments made at the very beginning of the trial and were not views that were shared after embedding FLASH within the curriculum.

Wider pupil outcomes

In addition to perceived direct effects on attainment, many teachers from intervention schools reported that FLASH had positive effects on other outcomes. In line with the logic model and implementation plan, some teachers highlighted improved engagement and motivation in Key Stage 4 English, particularly in relation to pupils' willingness and ability to write effectively and to meet the requirements of the exam specifications. Linked to this, teachers also noted improved metacognitive and self-regulation skills. These were identified by the development team from the outset as being important 'ingredients' for contributing to improved attainment in English. Heads of Department noted an improvement in pupils' levels of independence with marking and assessing their own work or that of their peers. Increased 'ownership' and 'responsibility' were phrases that came up regularly in our discussions and via correspondence with teachers.

> *Being a part of the FLASH project has pushed us forward departmentally with training our students to be metacognitive. We have an understanding and route-map to facilitating self-regulated learning. Our students across key stages now take ownership of their own work and improvement processes.*
>
> (Head of Department, Shelley Hub, South-East)

> *Flash Marking has really helped pupils develop their metacognition, especially using it throughout the planning, writing, reflection stages of the work cycle. They are much more aware of and able to control and articulate the way they think about their own work and skills development.*
>
> (Head of Department, Dickens Hub, South-East)

In two case study schools, metacognition and pupils' ability to understand and examine their own work were key outcomes of the intervention. In a case study school in the Shakespeare Hub, where there was considerable variation across the department in terms of implementation, one teacher described FLASH as 'most successful as a metacognition tool for the kids' as it meant that the pupils could identify areas for development more effectively. A Head of Department from the Wordsworth Hub also made the following comment:

> *For myself, the main success of this project has been seeing our students' confidence increase alongside their understanding of the skills that they need to master. I love listening to students debate whether a word that has been used is 'ambitious', or whether they have analysed a quote or just explained it.*
>
> (Head of Department, Wordsworth Hub, North-West)

Our discussions with pupils reflect positive perceptions about these 'softer' outcomes. Pupils tended to like FLASH Marking as an approach and in our discussions in case study schools, they often explained that the codes encouraged

them to think and reflect on their work. Again, some commented on how the colour-coding made this a visual and more concrete strategy for identifying improvements. Teachers that we observed, particularly during the earlier visits, were sometimes quite explicit with sharing the benefits of FLASH Marking with pupils, including how it provided personalised feedback and helped pupils to know how to improve. Pupils' views on this were exemplified by comments such as the comment below, taken from the pupil questionnaire:

> *Our English teacher sometimes makes us self-assess our assessments first using flash marking and then marks them himself after we've reflected on it ourselves, which I think is useful because we can see ourselves what we think we need to do to get better and then compare it to what the feedback we get afterwards.*
>
> (Year 11 Pupil, Dickens Hub, London)

Teacher workload

In the second teacher survey, 76% (n=121/159) of intervention group teachers reported that FLASH Marking had reduced their marking and feedback workload. Just 15% disagreed that it had this impact. These figures align with the analysis of perceived impact conducted above, which indicate that FLASH Marking did have some effect on reducing the amount of time that teachers were spending on marking and feedback activities. A number of Heads of Department commented further on the impact of FLASH in the workload survey.

> *There is no doubting that it has reduced teacher workload when it comes to feedback and assessment, allowing us to spend more time on the important things like planning. The department were a little reluctant to change at first, but once they saw it working they fully bought into it and have reaped the benefits of the speedy nature of the initiative.*
>
> (Head of Department, Dickens Hub, London)

> *I think it has taken a while for every teacher to fully get on board. Early adopters are ahead in their journey as a consequence. But when the impact it has had on the marking workload became apparent and people began to see the impact that metacognition has on quality of work when they took on classes that had been FLASH trained after Year 1 of the trial, its spread is now complete as our way of seeing assessment.*
>
> (Head of Department, Shelley Hub, South-East)

Interestingly, both of these comments also draw upon issues of staff 'buy-in' to the intervention, suggesting that once evidence of a reduction in workload became clear, teachers were more likely to engage and implement as expected. The subsequent impact of increased time for planning is also noteworthy from the first comment, as is the view that improved metacognitive skills have improved pupils' academic work, as mentioned in the second comment. These perspectives are indicative of how English teachers often discuss FLASH Marking by intertwining the various 'key ingredients' and outcomes involved. This teacher summarised how the dual proposed outcomes of pupils' progress and staff workload appeared to be working in tandem:

> *We can all agree that our workload has definitely reduced. Marking a set of books is no longer such a daunting process but the students' responses to feedback is now to a much higher standard and they are taking more ownership and pride in their reflection time.*
>
> (Teacher, Wordsworth Hub, North-West)

Our visits to case study schools and discussions with English teachers, Heads of Department, and senior leaders reflected similar views to those described above. Where FLASH had been well embedded, teachers typically felt that they could detect some reduction in marking workload. Where the intervention was still being used alongside other methods (particularly longer written comments), teachers unsurprisingly, felt that FLASH Marking was not reducing workload, and in some cases was increasing it. We encountered this view in three case study schools, with reference to a small number of English teachers in each department. In one case study school, the FLASH Marking Lead teacher

reported that some of her department still felt 'guilty' about using FLASH as opposed to writing longer comments. As such, they chose to ignore some of the FLASH guidance and opted to regularly write extended feedback on pupils' work; when discussing FLASH Marking in department meetings, they reported being aware that this approach took more time, but they felt that it was an essential part of their role.

The stage of the trial also made a difference to perspectives on this issue too. Earlier in the evaluation, teachers were more likely to report that FLASH Marking was time-consuming. This was due to the requirements to attend training, examine and amend schemes of work and curriculum planning, to become familiar with the new approach, and spend time discussing and using it with pupils. All of this required a considerable investment of time in the initial months, although there was a feeling that this was 'repaid' once the intervention was embedded, and the workload reduction became evident.

Contributing to the view that FLASH Marking could reduce overall marking/feedback workload, many teachers also reported that FLASH Marking allowed for a quicker turnaround of marking pupils' work. In a number of our case study schools, teachers suggested that they could assess a set of books much more efficiently than if they were conducting fuller, comment-based marking. Some teachers explained this by saying that if pupils were conducting self-/peer-assessment with FLASH Marking during lessons, then teacher assessment following this was considerably quicker as they were not having to identify and repeat the same feedback as had already been noted. Most felt that this was beneficial for gaining a swift overview of how a class was performing and also for addressing issues more quickly on a whole-class basis, while still being able to provide personalised (code-based feedback) for individual pupils where needed.

Teacher practice

FLASH Marking was deliberately designed as an intervention, which could be adapted by schools to suit the particular needs of their staff and pupils. The developers felt that it was important that English departments could tailor the FLASH Marking codes and approaches, depending on the exam boards and specifications that they were working with, the level of expertise and experience of staff, and the characteristics of the pupils in their classes. For many departments who signed up to participate in FLASH Marking, this flexibility was viewed as a positive element.

As we note in a number of the sections above, teachers' who committed more fully to the intervention, tended to see a change in their general day-to-day teaching practice. Using FLASH Marking was regularly 'built in' to English lessons and was then used by teachers to mark and provide feedback on written work. Via the training (both from the development team and the cascaded training in schools), some teachers—in our case study schools—reported being introduced to a range of related research, evidence, and resources. Some were very enthusiastic about this professional development and new learning and remarked that it had influenced their wider practice too. We saw examples of teachers applying some of the training around metacognition and curriculum sequencing, which had been delivered by the development team, to non-FLASH Marking content and year groups. These areas were becoming a more embedded part of teachers' role, with the training providing a helpful 'springboard' to developing their practice more generally. One Head of Department explained that because the training had been delivered by representatives from a research school, they felt confident embracing and employing it on activities beyond the FLASH Marking evaluation. For this teacher, and some others that we spoke to, the broader principles of the intervention had played an important role in developing their practice and facilitating a new approach to teaching GCSE-level English. These issues are discussed further in the section on 'Sustainability and Scalability of FLASH Marking' below.

Sustainability and scalability of FLASH Marking

As described previously, staffing was an integral element of ensuring that FLASH Marking was implemented and sustained within a department. Where staff who received the original training left the school, momentum with the intervention often diminished, and effective use of FLASH Marking declined. In contrast, many schools in the trial did sustain their levels of engagement. Our visits and discussions with school colleagues and the development team showed that a number of schools did plan to continue using FLASH Marking after the end of the evaluation. By September 2019 (one year into the trial) 31/52 schools had rolled-out FLASH Marking to all of their Key Stage 3 year groups (Years 7, 8, and 9). A total of 35 schools were also using FLASH Marking with the Year 10 group in addition to the Year 11 cohort who were the focus of the evaluation. A total of 14 schools also reported that other subject departments (such as History, Modern Languages, Science, and Media Studies) had introduced FLASH Marking too. Others reported that they had plans to do this later in the year, although we are not sure whether this did in fact happen due to the disruption caused

by Covid-19. These figures indicate that, at least for some schools in the trial, there was a view that FLASH Marking is a sustainable and worthwhile approach, and that there are possibilities for use with different age groups and for subject-specific adaptation of the intervention. In spring 2020, one Head of Department shared these positive views:

> *We are now in the Final Countdown of this trial…and we've come a long way. We are all committed converts to this approach, as it has made us rethink our pedagogical practice, the structure of lessons, and the quantity and quality of teacher feedback. So much so, that we have rolled FLASH out to the current year 10s, and following the completion of this project, are planning to implement this in KS3 [Key Stage 3] too. I have also used it with KS5 [Key Stage 5] to help with essay writing skills.*
>
> (Head of Department, Wordsworth Hub, North-West).

In one of our most committed case study schools (in the Shakespeare Hub, Midlands), the Head of Department remarked that they would continue using FLASH Marking after the end of the trial, regardless of the evaluation outcome. She felt that the intervention had had such a positive impact on staff and pupils, and that this had been acknowledged by senior leaders who had recently highlighted practice and exercise books in English as among the best in the school. These comments and others like them that we heard, reflected the general consensus that once embedded effectively, FLASH Marking could continue to be implemented at relatively low time or financial cost to those involved. Many teachers could not imagine returning to a more traditional approach of lengthy comments and/or frequent alphanumerical grades.

The figures above indicate that not all schools planned to continue with FLASH Marking after the end of the trial. This information was collected by the development team, and we did not probe for more detail about these decisions. However, in contrast to the positive views exemplified above, one FLASH Marking Lead from a case study school in the Shakespeare Hub (Midlands) explained that they would not continue with the intervention because they found it more 'onerous' than their previous approach to marking and feedback. They acknowledged that staff engagement and commitment to the intervention in their school had been patchy overall (also reflected in our data from the visits across the two years of the trial), and that this had led to a view that using FLASH was 'more difficult' than methods they were more familiar with. With the data that we have, it is difficult to quantify how widespread this view was across different schools, and this is something that could be returned to, should a scaled-up version of FLASH be considered.

The generally positive perspectives above, suggest that there may be a rationale for the expansion and scaling-up of FLASH as an intervention. The training is the main external input for schools to engage with. If the intervention was to be scaled-up, then the development team may need to expand the number of trainers they had available in order to be accessible and available for a wider group of schools. This could also allow for additional coaching or support visits where staff turnover in schools occurs.

Usual practice in control schools

As part of the IPE, we collected data from a small number of control schools (n=3) using case study methods. We have also drawn upon information provided by teachers in control schools in the workload survey. These data provide important insights into the kinds of practices being used in non-FLASH schools, and of the kinds of concerns that teachers and leaders were raising in relation to marking, feedback, and teacher workload.

Comments in the second teacher survey indicated that workload remains an important and emotive issue for many teachers working in control schools. A number of teachers reported that their workload as English teachers was considerably higher than that of teachers from other subjects. Teachers also expressed concern about the variation in marking load across the term: at the beginning of the new term it can be lighter, but later in the term, when all year groups are completing assessments, it can be very high and feel very pressurised. A number of teachers also shared concerns that they were not doing enough for their pupils because they could not keep up with the level of marking required. Others remarked that marking expectations often meant that other elements of the role or teachers' personal lives, were neglected:

> *I constantly feel that I don't do enough marking and that my lessons aren't sufficiently informed by marking. I am busy outside of school and refuse to spend my 2 days off working all day, which I easily could and sometimes have to. As an experienced teacher, I make judgements about what I can get away with dropping and what is essential. However, I am left with a nagging frustration that students are not getting what they should. It's quite stressful*.

(Teacher, Control School, South-West)

> *The frequency of marking is far too much to be expected of our subject. Quite often, frequency of marking and the time spent on this can affect the quality of lesson planning. Mocks and assessments should not be placed together in the calendar for all year groups if we are not to be given extra time to mark. Eating into family time far too much.*
>
> (Teacher, Control School, London)

Two teachers noted more manageable approaches to marking in their schools; however, as the second comment indicates, there were still genuine concerns about workload more generally.

> *There is now less onus on teachers making detailed written comments, which is positive. We are starting to use marking codes and pre-typed criteria and assessment grids to avoid having to write out multiple comments.*
>
> (Teacher, Control School, South-East)

> Our school trusts its staff; we do not, currently, have a draconian policy requiring us to mark books within a certain time frame. However, the more I compare teaching to other careers, the more ridiculous it seems. There is never enough time to do things properly.
>
> (Teacher, Control School, South-West)

Visits to our three control schools also indicate that workload was a factor when determining marking and feedback approaches. The three control schools all adopted 'whole-class' feedback strategies in the second year of the trial. This is an approach that has gained growing momentum since the start of the trial (see for example, Macdonald, 2021; Riley, 2020), and has been particularly championed by English teaching practitioners. Whole-class feedback can encompass a range of different approaches, but the aim is to summarise key issues and areas for pupils to develop on a class level, providing helpful targets for all learners, while reducing the in-depth marking of individual pupil work. Our case study schools were employing whole-class feedback to varying degrees. For one, it was the predominant method of feedback on pupils' written work; for the other two, it was used in conjunction with other approaches such as 'maintenance marking' (i.e. ticking to acknowledge that work had been seen/read) or code-based strategies used by teachers/pupils. The teachers at the three schools spoke positively about whole-class feedback for supporting pupils' learning and for promoting independence. At one of the schools (in the West Midlands) however, a teacher felt that when combined with marking of individual exercise books, it did not reduce the overall workload.

Our visits to the control schools demonstrated how the area of assessment and feedback was often in flux. As we note above, over the course of the trial there had been considerable developments to assessment policy and practice in these schools. These changes were sometimes aligned with wider issues such as curriculum re-design, or in one case, the strategic vision of a new leadership team. During our visits earlier in the trial, teachers reported shifts in the department and school-level assessment policies, again reflecting a general aim of reducing the marking burden for teachers while retaining high-academic standards. As some of our FLASH Marking teachers noted, however, the move to a new system sometimes brought with it additional challenges, including further time needed for familiarisation and embedding new approaches into daily practice. A teacher at a control school in the East Midlands commented, for example, that:

> *…there has been a redraft of the curriculum which has made things more challenging. The department now has zero capacity and we have no additional time for any sort of extra activities. Everyone has a full-time timetable and there is the implementation of the new KS3 [Key Stage 3] curriculum…there is now an emphasis on whole-class feedback for assessments and the department are launching a progress tracking system for the whole school too.*
>
> (Teacher, Control School, East Midlands)

In addition to the above, the school were also about to start their own school-level 'trial' of different marking and feedback methods. The above comments highlight the pressurised environments that teachers are working in, and the fact that even well-meaning approaches, which are designed to improve children's education and conditions for teachers often come with some cost in time. By the second year of the trial, two of the control case study schools had also recently moved to only providing in-depth feedback on 'assessed' pieces of work that were conducted once per half-term. This

was as a result of concerns about workload. With these pieces, teachers would also provide grades/marks to pupils, which would be tracked within the department and reported home to parents. In one of the schools (in the South-East), the Head of Department explained how time was used in class to consolidate the feedback given and the marks received. Pupils were supported to reflect upon their attainment to develop their own targets for future work. Verbal feedback had become a more central element in lessons, and the Head of Department explained that this meant pupils could receive and act on the feedback more quickly than with written comments. During a discussion with a small group of pupils at the school, they shared their views on the marking and feedback approaches being used.

> *Verbal feedback is the best. The teachers really try to give verbal feedback to all students, but there is not enough time to give enough verbal feedback.*
>
> (Pupil, Control School, South-East)

> *I like verbal feedback because it is more specific. Teachers also ask students who are really good at a specific area to give verbal feedback to their peers who ask for help. We really like this. You can ask your mate and they can explain to you something, sometimes better than teachers.*
>
> (Pupil, Control School, South-East)

One of the control schools was using a code-based approach to marking (along with the whole-class strategy described above). Codes were aligned to targets (for example, T1 – use quotation) and were written for pupils in their books. While in some ways similar to the FLASH approach, teachers at this school had not received the training associated with FLASH Marking, nor were the codes closely aligned to GCSE exam board specifications. There was some evidence of colour-coding alongside the use of the code-based feedback, which is also somewhat similar to the FLASH approach used in some schools.

Overall, though, the control schools that we engaged with as case studies did not appear to be consistently using approaches that were similar to FLASH Marking. While there seemed some acknowledgement of workload issues, and some attempt to reduce workload via newly implemented marking policies, the strategies that we saw did not encompass the multiple different elements that can be observed with FLASH. One teacher who responded to the survey reported the interesting experience of moving from an intervention school to a control school during the course of the trial. Her perspective is helpful for understanding the different approaches used, and the challenges associated with dismissing a strategy that she felt was effective (in this case, FLASH):

> *I have had the benefit of moving schools—both of which have been involved within this trial. I started at [School, Dickens Hub, South-East] and when my then HOD [Head of Department] said we were taking part in a Flash Marking trial I was intrigued. It turned out—I was already using a variety of codes in my marking to aid pupil understanding of the success criteria. The Flash Marking system which [School] took part in just gave a degree of rigour and formality to what I was using. I then moved to [Control School]…This was difficult as I had already been used to marking with codes…I did try to control marking and not use codes but once you start using them and the kids understand how to use them it really is difficult to not use them. They are easy to model, scaffold and identify within exempla answers. What's more you can track them across a student's work and use this data to hone in and develop writing more so.*
>
> (Teacher, Moved from Dickens Hub, London to another school in the South-East)

While we have noted some potential overlap with marking/feedback approaches (in the comment above and in the code-based approach used at one of the control schools), we think the likelihood of widespread contamination seems low. Control schools were not able to access the training or resources associated with FLASH during the course of the trial, and from our data, it seems that most either continued with their pre-trial practices or undertook some assessment policy amendments as a way to reduce workload. There is no evidence of alternative strategies being used with the same level of focus on teaching, learning, and professional development as we see with the FLASH intervention.

Cost

This section discusses average costs of delivery of FLASH Marking over the course of three years. The costs noted in Table 19 below are those, which a school would need to pay if they chose to 'buy-in' the FLASH Marking intervention.

These are different to costs incurred by schools during the trial as intervention schools did not have to pay the trainers' fees nor travel and accommodation. Moreover, costs for cover for teachers attending the training was paid to schools via the incentive that they received for participating in the trial. Schools did not have to purchase a licence for the online Trello portal and all printing costs were covered by the development team.

In line with the EEF guidance, costs for resources that schools would already be expected to have (for example, teachers, highlighter pens, and exercise books) are not included. For future roll-out or scale-up of the trial, however, we would recommend that some cost is directly allocated to purchasing highlighter pens. As we have shown above, the highlighting of written work emerged as a core feature of the FLASH Marking approach. This was an area that both staff and pupils were very positive about and it appeared to be clearly relevant for objectives linked to code use and pupils' development of metacognitive skills via self-/peer-assessment. Budgeting for these items would therefore, seem helpful for supporting effective implementation of the intervention and pupil engagement.

The costs below are based upon figures provided by the development team in 2020–2021. Per pupil costs are based upon the average cohort size included within the trial (n=183 pupils) and a single Year 10 cohort of pupils engaging with the intervention for the two years of their GCSE programme. Costs would naturally be reduced if a school decided to roll-out FLASH Marking to other year groups (as has happened with a number of schools involved in this trial). The EEF's cost rating can be found in Appendix A.

Table 19: Cost of delivering FLASH Marking

Item	Type of cost	Cost	Total cost over three years	Total cost per pupil per year over three years
Trainer fees x three days training x two trainers	Start-up cost per school	£350 per day	three days = £1,050	£5.73
Trainer travel and accommodation x three days x two trainers	Start-up cost per school	£400 per day	three days = £1,200	£6.55
Supply cover for teacher absence to attend training	Start-up cost per school	£150 per day per teacher	three days x two teachers x £150 = £900	£4.92
School licence for online portal	Start-up cost per schools	£50	£50	£0.27
Printing of training resources	Start-up cost per school	£50	£50	£0.27
Trainer fidelity and coaching visit to school and travel / accommodation	Ongoing school cost (one per year)	£175 (fee) plus £200 (travel and accommodation)	£375 x three days = £1,125	£6.15
Total			£4,375	£23.89

Table 20: Cumulative costs of FLASH Marking (assuming delivery over three years)

	Year 1	Year 2	Year 3
FLASH Marking	£3,375	£1,125	£1,125

In response to the Covid-19 pandemic, the development team significantly altered the FLASH training to reflect the challenges faced by English departments across the country. This newer, online course delivered via Zoom, allows English teachers to access the intervention training at a lower cost than that, which is specified above. Travel/accommodation costs and printing of training resources are not necessary, and the newer version of the programme does not involve the use of the online Trello platform either.

Conclusion

Table 21: Key conclusions

Key Conclusions
1. Due to Covid-19 the primary outcome for this evaluation (GCSEs) was not collected and so no measure of impact on English Language and Literature attainment is reported and there is no security rating for this trial. Key conclusions are based on qualitative data from the implementation and process evaluation.
2. Teachers receiving FLASH Marking reported a greater reduction in both total hours spent working and hours spent marking compared to teachers who did not receive the intervention. This was reported by teachers who completed both workload surveys—pre- and post-intervention—due to dropout, this was approximately 25% of the number of teachers who responded to the initial survey.
3. Process evaluation findings indicate that attendance at training was high, and the intervention was generally received positively by teachers. FLASH Marking was implemented as intended in most schools with some variation, mostly due to staff engagement. In total, 90% of English department Heads responding (n=40) to the second workload survey strongly agreed or tended to agree that their department had been fully committed to FLASH Marking since the start of the trial.
4. As observed in training sessions, case studies, and staff questionnaires, where leadership personnel changed, a school's emphasis on and engagement with FLASH Marking sometimes changed too. Where senior leaders were also members of the English department, there tended to be increased buy-in to FLASH Marking.
5. Teachers were pleased with the quality of training and the support available from the delivery team. Discussions of good practice across school hub groups were received enthusiastically, particularly in the earlier stages of the trial. 14 schools reported that they had cascaded the FLASH Marking training to other subject departments in their schools.

Impact evaluation and IPE integration

Interpretation

This evaluation was the first trial of the FLASH Marking intervention. It was designed as a relatively large-scale efficacy trial, involving 103 schools across England and nearly 18,500 pupils, and was due to run for two years between 2018–2020. The original aim of the evaluation was to understand whether this new approach to marking and feedback—using codes, reduced grading, and incorporating high-quality CPD for teachers—could lead to improved academic outcomes in GCSE English and English Literature. A secondary outcome focused on whether FLASH Marking would have an impact on teachers' workload. In addition to this, the impact evaluation was designed to be supplemented by a rich, in-depth IPE. This involved working with a small number of schools over the period of the trial in order to gather detailed information about how FLASH was working from the perspectives of teachers and pupils, and to contribute to our explanations of the impact evaluation outcomes. As we note above, due to Covid-19, we were not able to complete the trial in the way that had been planned; GCSE grades were awarded via teacher assessment rather than through external examinations. As such, we were not able to conduct the analyses for the primary outcome (attainment). Instead, this report presents findings from the teacher workload outcome from the project, and also summarises the extensive IPE work that was undertaken. Our findings from the analysis of these areas suggest that FLASH Marking is a potentially effective intervention, which seems popular with teachers. As a result, there may be value in conducting a further study in order to understand its impact on children's academic attainment. We summarise our key conclusions, alongside existing evidence in this area below.

Our evaluation found that English teachers in FLASH Marking schools reported a reduced number of working hours compared with teachers working in control schools. The post-test survey conducted 16 months in to the FLASH evaluation, showed that FLASH Marking teachers reported an average working week of 41.01 hours, and those in control schools 45.32 hours. When considering marking and feedback activities specifically, the FLASH group also reported a lower average number of hours: 5.69 hours for the intervention group and 7.05 hours for the control, and an 'effect' size of -0.27. Despite this seemingly positive effect, we interpret these figures with caution. Sampling variation is an issue here as both groups include many teachers who either left their original school (where they completed the first survey) and/or had new teachers joining part-way through the trial. Moreover, the response rate for the second survey was lower than intended (n=358 responses from a total of approximately 950 teachers still involved in the trial by spring 2020). We discuss possible reasons for this below (Limitations and Lessons Learned section) and consider potential strategies to improve this if a similar approach were to be used in future studies. The sampling variation across the two surveys should be considered when interpreting the findings here as it is unlikely that staff turnover occurs 'randomly', with some

schools perhaps more likely to have higher teacher attrition than others. This is also true of teachers choosing not to participate in the second survey; it is possible, for example, that those who did not respond, already had a higher workload and thus did not have time or inclination to participate. A further issue lies in the fact that teachers in the intervention arm were not (and could not be) blind to receiving the intervention. As such, findings from reported workload data could be attributable to social desirability effects, for instance, participants responding in a way that they believe may be perceived positively by others. We cannot be sure of the exact nature or impact of these sampling issues but, as a result, we do urge caution in the interpretation of the findings.

In order to provide a more rigorous measure of workload variation over the period of the trial, we also matched the responses from the teachers who had responded to both surveys. On both measures (overall workload and marking/feedback workload), respondent teachers in the intervention group reported lower numbers of hours compared to the control group, with 'effect' sizes of -0.16 and -0.17, respectively. These findings also suggest that FLASH Marking could have had a potentially positive impact on the workload of those teachers involved in the trial. It should be noted that the workload for the control group also declined over the course of the evaluation (although not as much as the intervention group), suggesting that these schools might also have been adapting their practice to try and address teacher workload (as discussed in a review of the practices used in a small number of schools by Elliott *et al.*, 2020).

Overall, the schools and teachers that we engaged with as part of the process evaluation were positive about FLASH Marking. Perspectives collated from Heads of English and English teachers align with our findings in relation to workload. The majority of teachers that we spoke to felt that FLASH Marking did contribute to an overall reduction in hours worked, particularly in relation to marking and feedback. They acknowledged that time was needed at the outset of the trial to embed the intervention effectively, but once this had been done, less time was required for providing written feedback. Compliance with attendance at training and cascading of training to departments was strong. Teachers also reported being satisfied with the quality and quantity of support provided by the development team.

Despite these generally positive views, teachers and pupils did share some areas for development of the intervention. One of the concerns centred on the number of codes, with some teachers and pupils finding that there were too many to fully understand and utilise. Some teachers adapted the intervention to reduce the number of codes used. According to the development team, this was an acceptable amendment, and one which they encouraged schools to do in order to ensure that FLASH was tailored to the needs of the department and pupils. Heads of Departments in two case study schools described introducing codes in small 'sets' to pupils so that they were not overwhelmed by the new information. A teacher in a third school, used short retrieval activities to remind pupils of what each code meant. These kinds of approaches, combined with helpful resources such as classroom displays, could be useful for supporting pupils to become familiar and confident with the range of codes and their meanings.

In some schools, we saw some variation in the extent to which FLASH was being implemented. As we note above, we saw some evidence of 'intervention drift' in the second year of the trial. Some schools appeared to be using grades/marks more frequently than they had done in the first year of the trial. Two Heads of Departments explained that this was due to the pupils being in Year 11, their final year of school, and the year when they would complete their GCSEs. They felt that it was important for the pupils to know 'where they were' in terms of their academic grades and progress. In one case, there was also a wider school expectation that pupils were regularly informed of their marks too. This shift between the first and second year of the trial is potentially important as it may have implications for the longer term sustainability and successful implementation of FLASH Marking in a Key Stage 4 context. It may also indicate that additional input might be required in the second half of the trial to support departments with ongoing delivery of FLASH as intended. Staff turnover was also a key factor in determining variation of FLASH implementation; teachers who had not been in the school at the start of the trial appeared less likely to engage with the intervention and to embed it within their practice. The development team aimed to provide additional support for teachers who joined the trial part-way through, but this is something that could be tackled further with additional resources. In some schools, teachers also commented on the value of colour-coding the work (using highlighters) but pointed out the costs involved with regular use and supply of these. Finally, we heard some mixed views about how FLASH Marking was used with and perceived by groups with different levels of prior attainment. Some teachers felt that FLASH particularly suited those with lower prior attainment while others felt that it was less suitable and potentially confusing for these pupils. It would be useful to gain further information on this and to consider how this might feed-in to future development of the intervention and training with teachers.

The FLASH Marking evaluation was designed to generate rigorous evidence on the effectiveness of approaches to marking and feedback in schools. A review by Elliott *et al.* (2016) pointed to the lack of high-quality research examining the impact of written marking on pupils' attainment. Unfortunately, due to the Covid-19 pandemic, this evaluation has not been able to contribute to this gap in our knowledge, although we do feel that there is potential for this in future. FLASH Marking aligns closely with the EEF's guidance on the principles and methods for providing effective feedback (Collin and Quigley, 2021). Moreover, there are elements of the intervention, which connect with other recent, successful evaluations involving formative assessment. The Embedding Formative Assessment approach, for example, included high-quality professional development and teacher networks as part of its delivery (Speckesser *et al.*, 2018). There is also some promising evidence from feedback approaches, which include strategies for pupils to develop their metacognitive skills (Motteram *et al.*, 2016; Muijs and Bokhove, 2020), although again, there is very little work, which explicitly examines this in relation to written marking/feedback (Elliott *et al.*, 2016; Newman *et al.*, 2021). In light of this, we feel that there is merit in carrying out further evaluation work, which considers the role of written marking/feedback on young people's attainment; this is also in line with the EEF's current funding priorities (EEF, 2021b).

Evidence to support the logic model

Without attainment outcomes, it is difficult to be conclusive about the extent to which this study has aligned with the anticipated theory of change. As we have highlighted in our process evaluation, we feel that there would be value in examining more closely the extent to which high fidelity to the intervention impacts both attainment and workload outcomes. It is possible, for example, that medium-low fidelity (as we saw in a small number of schools) may still be enough to contribute to positive outcomes in one area, such as workload, but less so for pupils' attainment. Having a clearer picture of this issue would enable us to draw more practical conclusions about, which 'ingredients' or mechanisms are likely to be contributing to the selected outcomes.

The results from this evaluation provide some support for the original logic model and offer some evidence about key aspects of the implementation (see above but also below for limitations in Limitations and Lessons Learned section). In relation to teacher workload, our evaluation points to mechanisms such as code use, reduction in comment marking and grading, and an emphasis on pupils independently reflecting on and assessing their work, potentially contributing to a positive outcome. Our IPE data indicates that the training may have been particularly influential here. These sessions were where representatives from each participating school learned about FLASH Marking, its implementation its potential impact. Attendance at the sessions was good and teachers shared positive views about the content and delivery. Fidelity measures carried out by the development team showed strong teacher engagement with the intervention, including regular use of the codes, embedding of FLASH Marking in to planning and assessment, and more regular feedback being provided to pupils. It is not clear that these are the direct mechanisms, which may have supported with some reduction in workload, but it is possible that they contributed in some way. From our IPE, there is some evidence that FLASH Marking supported a number of schools to embed self- and peer-assessment more meaningfully into English lessons, promoting pupils' levels of independence, and also perhaps contributing to some reduction in teachers' marking load.

There is limited evidence regarding the additional support that some schools received with FLASH Marking implementation, and its impact on teacher/pupil outcomes. The aim from the development team was to visit each school once across the course of the trial, both for fidelity purposes and to offer additional support or professional development where needed. There was some variation in this provision with some schools receiving more than one visit from the development team (n=9), most schools receiving one visit (n=23) and 20 schools receiving no visits over the course of the trial, up until March 2020. Despite frequent communication and offers of visits from the development team gaining access to a number of the trial schools was difficult. Five of the schools had also scheduled visits for spring 2020 but due to the Covid-19 pandemic these had to be cancelled. Schools who did have visits tended to be very positive about this input and welcomed the support and expertise of the development team. However, with the available evidence it is not possible to make judgements about the impact on subsequent practice and on the outcomes of workload or attainment. This is an element of the trial that would benefit from further scrutiny and development in future work, in order to better understand the purpose, value, and efficiency of in-person visits to schools.

Due to no primary outcome being collected for this evaluation, we have not included a revised version of the theory of change.

Limitations and lessons learned

The FLASH Marking evaluation progressed successfully for the first 16 months (September 2018–February 2019). For a relatively long trial involving a large number of schools, retention and engagement with the intervention were strong. With the arrival of the Covid-19 pandemic, however, the trial was compromised both because of school closures and the shift to online practice, and the cancellation of external examinations. While many schools continued to use FLASH Marking during the final stages of the trial, in reality this was made very difficult due to the circumstances. Teachers had to re-prioritise their work and support pupils through distance-learning approaches; these are not the conditions that FLASH was devised to be conducted in and we recognise that it would not have been feasible for many teachers to deliver the intervention in this way.

Plans were put in place for an extension to the evaluation. The EEF and the evaluation team were hopeful that we could use GCSE data from 2021, given that the majority of schools in the trial were using the intervention with that cohort of pupils too. While these data would not be as rigorous as those from the previous year and would require careful interpretation, this was a plausible alternative, which would yield some valuable information on the attainment of pupils based in schools that were using FLASH Marking. However, once the external exams were cancelled for a second year, the decision was taken to abandon the primary outcome measure and to conclude the evaluation with the existing data on workload and from the IPE.

One of the possible lessons to be learned from this trial is the value of having an interim measure of attainment. This recommendation emerges partly as a result of the Covid-19 pandemic and the disruption to the second part of the trial. In addition to this though, it is also a recognition of the length of the FLASH Marking trial and questions around 'how much' of the intervention is required to have an influence on attainment. It would be helpful to know, for example, whether one year of FLASH Marking is enough to make a difference to pupils' grades. This also would have been helpful for capturing outcomes from schools who engaged more fully with the evaluation in the first year. Having a short assessment of pupils' achievement in English (perhaps in the style of a mock exam) halfway through the trial would have provided an indication of any emerging differences between the intervention and control groups after one academic year of the evaluation. Either a teacher-assessed piece of work could have been used or the evaluation could have built in an externally marked assessment. Both, of course, may come with additional financial or time costs but arguably, would have been worthwhile for providing some interim measure of attainment and for furthering our understanding of how the intervention was working and being implemented one year into the project.

The IPE analysis discussed is based on some in-depth, longitudinal data collection with a small number of case study schools. We acknowledge that a number of these schools, although certainly not all, were engaged with the trial and enthusiastic about being included as a case study in the evaluation. Teachers at many of the case study schools were supportive of our data collection endeavours and were keen to share their experiences and views of the FLASH Marking intervention. This could have implications for the interpretation of IPE findings, as it is possible that reports from case study teachers and engagement with the intervention/trial were more positive than that of the intervention school cohort as a whole. Having this though, we also note that there was considerable variation in the levels of engagement and commitment, to both the FLASH Marking intervention and the evaluation, from case study schools, as evidenced by the attrition of two schools and the challenges of arranging visits with some others (see IPE Results sections above).

A final limitation of our study is the response rate to the second teacher survey. The lower number of participant teachers involved in this element of the study makes it more difficult to draw conclusions about the impact of FLASH on teacher workload. One of the issues here related to teacher turnover, which made matching responses between the first and second survey difficult. The timing of the survey also may have been problematic as it occurred during a busy spring term. While teachers were given a number of weeks to complete the survey and it was designed to only take 10–15 minutes to complete, we still received a small number of comments about it being an additional burden on their time. Covid-19 and the associated disruption to schools was also just becoming more evident at this point too, and this is likely to have influenced teachers' willingness or ability to complete the survey. We would suggest that, as with the first survey, these matters all need to be taken into account for future work in this area, but also that the offer of an incentive may be helpful for encouraging completion. At the outset of the project, schools were not able to be randomised or receive their incentive payment until English departments had completed the survey; holding a further incentive payment

to be paid to both intervention and control schools after the second survey could help with supporting participation in that phase of the study too.

Future research and publications

Further to the findings discussed above, future research urgently needs to focus on the impact of written marking/feedback on young people's attainment. Despite having a heavy presence in schools and considerable implications for the practice and workload of teachers, this remains an area with very limited evidence. It would be possible to conduct follow-up analyses from this study, either focusing on the teacher-assessed grades of pupils in the original cohorts or examining the attainment of other cohorts of pupils who have attended FLASH schools. Following the extensive planning and roll-out phases of the intervention and the investment in time and resources from schools and teachers, we argue that obtaining some kind of attainment outcome for this trial would be beneficial and would inform other possible studies in this area.

Given the enthusiasm for many of the FLASH Marking schools to use the intervention with their Key Stage 3 (age 11–14) classes, it would be relevant to examine the implementation and impact with this alternative age group. Many of the teachers in the study felt that FLASH worked very well with these younger year groups, and so it may be possible to trial its feasibility and impact on attainment/workload with these pupils. This may also be viewed by schools as a 'less risky' option than trialling a new approach with GCSE cohorts, and yet would provide valuable insights in the effectiveness of the intervention and the principles underlying it.

In addition to the code-based approach seen with FLASH, there is also scope for examining alternative strategies, which are becoming more widespread in schools. Whole-class feedback is one of these and was observed in all of our case study control schools as well as in some of the FLASH Marking schools. In line with the EEF's work on school choices, we suggest that helping teachers to understand the effectiveness of different approaches to written/marking feedback is vital and will empower school leaders and teachers to make informed decisions about policy and practice in their settings. These questions also led us to think further about what a 'control' group or 'business as usual' group involves within a trial such as this. Due to the FLASH Marking intervention being new and relatively unknown at the start of the trial, having a firm and concrete idea of what the intervention group would receive involved a degree of uncertainty, and in turn, control schools were asked to continue with their 'business as usual' approaches to marking and feedback. Over the course of the trial, we have come to reflect that where practices, such as marking and feedback, are well embedded within schools, it can be challenging to understand and make sense of this group as a comparison cohort. As our small sample of case study control schools demonstrate, there was considerable variation *between* those in the control group. There is also indicative evidence that some of the control schools were employing approaches, which may have aligned or overlapped with some elements of FLASH Marking, for example, the use of codes, or the reduction of individual grading. However, we are limited in our understanding of these practices across the whole control group, making it difficult to conclude the impact of these issues. It is possible though that the impact of FLASH may have been downplayed as the 'business as usual' schools were all practicing feedback and marking (in some form) and so the difference between intervention/control schools is likely to be much less than if, for example, control schools were doing *no* marking or feedback. We are not suggesting that this should have been the case, but it is important to think about the effect of utilising a 'business as usual' group such as the one in the FLASH trial, and the implications for measuring overall outcomes and implementation. This is an area that we feel warrants further study and development if we are to fully grasp the complex causal mechanisms at play in an evaluation of this kind (see Perry and Morris, forthcoming, for fuller discussion of this issue).

Finally, this evaluation has provided further important data on the teacher workload situation in England. The responses to both surveys indicate the scale of the issue, particularly for English teachers. This report has highlighted findings relevant to the research questions outlined at the start of the project, but in addition, with the breadth and richness of the data yielded from the survey, there is certainly a case for sharing this information more widely and engaging in the debates and discussions, which are focusing on workload and its associations with teacher conditions, satisfaction, and retention. Our findings indicate that there is potential promise for FLASH Marking as a tool for reducing teacher workload, and as tool for offering a well-supported and evidence-informed approach to marking and feedback in schools alongside valuable professional development. As a result of this, we hope to disseminate findings from this study via both academic channels (for example, journal articles) and through practitioner-focused networks and publications.

References

Black, P. and Wiliam, D. (1998) 'Assessment and Classroom Learning'. *Assessment in Education: Principles, Policy and Practice*, 5: 1, 7–74.

Braun, V. and Clarke, V. (2006) 'Using Thematic Aanalysis in Psychology'., *Qualitative Research in Psychology,* 3: 2, 77–101.

Churches, R. (2020) 'Supporting Teachers Through The School Workload Reduction Toolkit', Education Development Trust. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/899756/Supporting_teachers_through_the_school_workload_reduction_toolkit_March_2020.pdf (accessed 15 April 2022).

Collin, J. and Quigley, A. (2021) Teacher Feedback to Improve Pupil Learning: Guidance Report, EEF. Available at: https://d2tic4wvo1iusb.cloudfront.net/guidance-reports/feedback/Teacher_Feedback_to_Improve_Pupil_Learning.pdf (accessed 15 April 2022).

Creswell, J. and Plano Clark, V. (2011*)* Designing and Conducting Mixed Methods Research, SAGE.

DfE. (2016) Teacher Workload Survey 2016, DfE. Available at: https://www.gov.uk/government/publications/teacher-workload-survey-2016 (accessed 15 April 2022).

DfE. (2018a) Schools, Pupils and their Characteristics, DfE. Available at: https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics (accessed 15 April 2022)

DfE. (2018b) Revised GCSE and equivalent results in England: 2016 to 2017, DfE. Available at: https://www.gov.uk/government/statistics/revised-gcse-and-equivalent-results-in-england-2016-to-2017 (accessed 15 April 2022).

DfE. (2019) Find and Compare Schools in England (Performance Tables), DfE. Available at: https://www.gov.uk/school-performance-tables (accessed 15 April 2022).

EEF (2018) Putting Evidence to Work: A School's Guide to Implementation, EEF. Available at: https://dera.ioe.ac.uk/31088/1/EEF-Implementation-Guidance-Report.pdf (accessed 15 April 2022).

EEF. (2021a) Teacher Feedback to Improve Pupil Learning: Six Recommendations for Using Teacher Feedback to Improve Learning, EEF. Available at: https://educationendowmentfoundation.org.uk/education-evidence/guidance-reports/feedback?gclid=CjwKCAjwkvWKBhB4EiwA-GHjFvQo96DPIuViOjAkv1kmrBq2muhEsL-TyUUqegOtMGcMk7HpuFoq-xoCjdYQAvD_BwE (accessed 15 April 2022).

EEF. (2021b) Assessment and Feedback: Improving the Teaching and Learning of assessment and feedback, EEF. Available at: https://educationendowmentfoundation.org.uk/guidance-for-teachers/assessment-feedback (accessed 15 April 2022).

Elliott, V. Baird, JA. Hopfenbeck, TN. Ingram, J. Thompson, I. Usher, N. Zantout, M. Richardson, J. and Coleman, R. (2016). A Marked Improvement? A Review of The Evidence on Written Marking, EFF. Available at: https://educationendowmentfoundation.org.uk/education-evidence/evidence-reviews/written-marking?msclkid=51da6868bcbc11eca409735e7e3ea59a (accessed 15 April 2022).

Elliott, V. Randhawa, A. Ingram, J. Nelson-Addy, L. Griffin, C. and Baird, J-A. (2020) Feedback in Action: A Review of Practice in English Schools, EEF. Available at: https://d2tic4wvo1iusb.cloudfront.net/documents/guidance/EEF_Feedback_Practice_Review.pdf (accessed 15 April 2022).

Gorard, S. (2021) How to Make Sense of Statistics, SAGE.

Hattie, J. and Timperley, H. (2007) 'The Power of Feedback'. *Review of Educational Research*, 77: 1, 81–112.

Independent Workload Review Group (2016) Eliminating Unnecessary Workload Around Marking, DfE. Available at: https://www.gov.uk/government/publications/reducing-teacher-workload-marking-policy-review-group-report (accessed 15 April 2022).

Information Commissioner's Office (no date) Special Category Data, ICO. Available at: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/special-category-data/ (accessed 15 April 2022).

Kime, S. (2018) Reducing Teacher Workload: The 'Rebalancing Feedback' Trial: Research Report, Evidence Based Education. Available at: https://dera.ioe.ac.uk/31210/1/Cheshire_Vale_-_Reducing_teacher_workload.pdf (accessed 15 April 2022).

Klapp, A. (2015) 'Does Grading Affect Educational Attainment? A Longitudinal Study'. *Assessment in Education: Principles, Policy and Practice,* 22: 3, 302–323.

Lipnevich, AA. and Smith, JK. (2009) 'Effects of Differential Feedback on Students' Examination Performance'. *Journal of Experimental Psychology: Applied,* 15: 4, 319–333.

McDonald, R. (2021) 'What Does Effective Whole-Class Feedback Look Like?' *Impact – Journal of the Chartered College of Teaching.* Available at: https://impact.chartered.college/article/effective-whole-class-feedback-english/ (accessed 15 April 2022).

Motteram, G. Choudry, S. Kalambouka, A. Hutcheson, G. and Barton, A. (2016) ReflectED. Evaluation Report and Executive Summary, EEF. Available at: https://www.research.manchester.ac.uk/portal/files/65367988/EEF_Project_Report_ReflectED.pdf?msclkid=cd7629f3bcbf11ec8ea5ccee4b0eac3c (accessed 15 April 2022).

Muijs, D. and Bokhove, C. (2020) Metacognition and Self-Regulation. Evidence Review, EEF. Available at: https://educationendowmentfoundation.org.uk/public/files/Metacognition_and_self-regulation_review.pdf (accessed 15 April 2022).

Newman, M. Kwan, I. Schucan Bird, K. and Hoo, H-T. (2021) The Impact of Feedback on Student Attainment: A Systematic Review, EEF. Available at: https://d2tic4wvo1iusb.cloudfront.net/documents/guidance/Systematic-Review-of-Feedback-EPPI-2021.pdf (accessed 15 April 2022).

Ofsted (2022) Find An Inspection Report, Ofsted. Available at: **https://reports.ofsted.gov.uk/** (accessed 15 April 2022)

Perry, T. and Morris, R. (forthcoming) Research Evidence and Educational Improvement: A Critical Guide Through a Divided Field*,* Open University Press.

Perryman, J. and Calvert, G., (2020) 'What Motivates People to Teach, and Why Do They Leave? Accountability, Performativity and Teacher Retention'. *British Journal of Educational Studies*, 68: 1, 3–23.

Riley, K. (2020) Whole Class Feedback: Reducing Workload, Amplifying Impact and Making Long-Term Change in The Learners, HoD and Heart. Available at: https://myhodandheart.wordpress.com/2020/10/09/whole-class-feedback-reducing-workload-amplifying-impact-and-making-long-term-change-in-the-learners (accessed 15 April 2022).

Speckesser, S. Runge, J. Foliano, F. Bursnall, M. Hudson-Sharp, N. Rolfe, H. and Anders, J. (2018) Embedding Formative Assessment: Evaluation Report and Executive Summary, EEF. Available at: https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/embedding-formative-assessment (accessed 15 April 2022).

Toropova, A. Myrberg, E. and Johansson, S. (2021) 'Teacher Job Satisfaction: The Importance of School Working Conditions and Teacher Characteristics'. *Educational Review*, 73: 1, 71–97.

Thomas, G. and Myers, K. (2015) The Anatomy of The Case Study, SAGE.

Zhang, B. and Misiak, J. (2015) ‘Evaluating three grading methods in middle school science classrooms’. *Journal of Baltic Science Education*, 14: 2, 207–207.

Appendix A: The EEF cost rating

Figure 9: 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

This project and its evaluation were affected by the 2020 partial school closures caused by the Covid-19 pandemic, and the cancellation of GCSE examinations in 2020. As a result, the evaluators were not able to use GCSEs in order to estimate the impact of the project on maths attainment. Therefore, it was not possible to rate the security of impact estimates.

Further appendices:

Appendices C to K are available as a separate document (Technical Notes).

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