Can a code-based approach to marking and feedback reduce teachers' workload? An evaluation of the FLASH marking intervention

Rebecca Morris, Stephen Gorard, Beng Huat See & Nadia Siddiqui

Published online: 10 Oct 2023.

Submit your article to this journal ⤤

Article views: 1564

View related articles ⤤

View Crossmark data ⤤

Routledge
Taylor & Francis Group

Can a code-based approach to marking and feedback reduce teachers' workload? An evaluation of the FLASH marking intervention

Rebecca Morris [a], Stephen Gorard [b], Beng Huat See [b] and Nadia Siddiqui [b]

aDepartment of Education Studies, University of Warwick, Coventry, UK; bSchool of Education, Durham University, Durham, UK

ABSTRACT

Teacher workload is an important policy concern in many education systems around the world, often considered a contributory factor in teacher attrition. One aspect of workload that could be addressed is reducing the amount of written marking and feedback that teachers do. This article reports on the results of an evaluation of FLASH Marking, an intervention aimed at reducing teachers' marking workload. FLASH Marking is a code-based feedback approach involving peer- and self-assessment, reducing the need to use alphanumeric grading while promoting the use of students' metacognitive skills. The study involved a single cohort of 18,500 Key Stage 4 pupils (aged 14/15 at the start of the trial) and their English teachers ($n = 990$) in 103 secondary schools in England. The impact of the intervention was estimated as the difference in before and after measures of teacher workload, comparing teachers in 52 intervention schools and those in 51 control schools. The results suggest that the intervention had the effect of lessening teachers' workload by reducing their working hours (effect size 0.16), including hours spent on marking and feedback (0.17). The intervention was largely implemented as designed and teachers were generally positive about the potential impact of FLASH on pupils' learning outcomes.

Introduction

Over the past 30 years, a growing body of research has pointed to the value of providing pupils with high-quality and timely formative feedback to improve progress and attainment (Black & Wiliam, 1998; Hattie & Timperley, 2007; Newman et al., 2021). When done well, feedback can help pupils to understand their strengths and areas for development, and identify where improvements are needed in order to move their learning forward. Despite recognition of the importance of feedback for learning, there is still an overall lack of strong research evidence examining the most effective approaches and methods that school leaders and teachers might choose to employ.

Recent reviews have particularly highlighted the limited number of large-scale, high-quality studies that consider the impact of written feedback on pupils' work (Elliott et al., 2016; Newman et al., 2021). Written feedback can come in different forms and may be used for different purposes, including provision of a mark/grade, sharing of targets or areas for development, acknowledgement that work has been completed, or praise. When providing written feedback, teachers are typically aiming to support and develop pupils' academic progress, although written feedback may also be used to achieve other outcomes such as improving students' motivation (Koenka et al., 2021) or metacognition skills (Motteram et al., 2016; Muijs & Bokhove, 2020). Elliott et al. (2016) note that written marking is just one form of feedback provision, but that it is often viewed as the main approach by which pupils receive feedback on their work. Written marking can include the use of grades, numbers, comments, or symbolic methods (for example, codes, images, and symbols), with schools often using different approaches and combinations of these.

Research on the impact of sharing grades for pupils' attainment has provided a largely mixed picture. A longitudinal study in Sweden, for example, compared the outcomes of 8,558 children in Grades 7 to 9, half of whom received attainment grades in Grade 6 (along with written comments) while half received only written comments (Klapp, 2015). This was a natural experiment where the introduction of a new curriculum allowed municipalities to decide whether to provide marks to children in Grade 6. The study reported that, when receiving both grades and comments, there was a positive long-term effect for girls, particularly higher-attainers. Boys and lower-attaining pupils generally did less well. 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. A recent meta-analysis (Koenka et al., 2021) which compares the impact of grades versus comments on students' academic performance finds that students who received comments on assessments performed nearly a third of a standard deviation higher than their grade-only peers. These findings and those of other smaller-scale school-based studies (e.g. Zhang & Misiak, 2015) or those conducted with college-level students (Lipnevich & Smith, 2009) all point to the value of detailed, performance-focused comments for improving attainment.

In both policy and practice, the terms 'marking' and 'feedback' often hold varied meanings for school leaders and teachers. As such, the implementation of feedback provision to children and young people in schools is extremely diverse and sometimes difficult to 'capture' accurately through research (Elliott et al., 2016, 2020). In addition, a set of wider concerns have come to the fore in the last decade, with practitioners and policymakers highlighting the problematic nature of excessive marking and feedback approaches for teachers' workload (Independent Workload Review Group, 2016). The heavy workload created by marking has been linked to reduced teacher job satisfaction and retention (Perryman & Calvert, 2020; Toropova et al., 2021). In England, there has been a growing interest from government, unions and teachers 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 designed to reduce the amount of written marking (e.g. Churches, 2020; Kime, 2018; Speckesser et al., 2018). There has also been development of strategies such as whole-class feedback as an alternative to more individualised feedback, and 'live'

marking, as a quicker approach to tackling pupil errors or misconceptions (Elliott et al., 2020; McDonald, 2021; Riley, 2020).

Within the diverse range of methods used for providing written marking, code-based approaches have historically been used for feedback on spelling, punctuation and grammar issues. 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. Some schools also use colour-based coding approaches (e.g. red, orange, green) to signify levels of understanding or goal achievement (Elliott et al., 2020). However, as far as we are aware, there have been no robust, school-based studies which examine the use of these code-based approaches nor the impact of more sophisticated code-based methods which support strategy or process (e.g. promoting pupils' self-assessment of their work, or supporting them to set targets and improve).

This paper presents a summary of an evaluation of one such approach known as FLASH Marking. This is described in the next section, followed by the methods used, and the process and impact findings. The paper ends with a summary of limitations, and discussion. Fuller details appear in the project report (Morris et al., 2022).

The intervention

FLASH (Fast Logical Aspirational Student Help) Marking is an approach to marking and feedback, developed by school leaders from a school in the north west of England in 2015/16. After some initial piloting with local schools, the Education Endowment Foundation (EEF) in England commissioned an evaluation of the approach in an efficacy trial. This was one of the largest trials of feedback approaches in schools conducted in England. The trial was designed to examine pupils' attainment in English Language and Literature the end of Year 11 (age 16), along with its impact on teacher workload. However, the Covid-19 pandemic led to externally assessed GCSE (General Certificate of Secondary Education) assessments being cancelled in 2020 and 2021, and thus data were not available to complete these analyses. As such, the intervention's impact on pupil attainment is not covered in this paper. Below we provide further details of the intervention and trial before presenting the findings relating to the effect on teachers' workload.

FLASH Marking is a teacher-developed feedback approach in which teachers use skills-based codes and brief comments rather than grades and fuller comments. The codes used in FLASH Marking are aligned with terminology and skills from the GCSE English Language and English Literature grade descriptions and are presented to pupils to signal where they have demonstrated certain skills and where there are areas for improvement and development. Figure 1 shows an example from a FLASH Marking code sheet, and the structure within which codes (or skill areas) were grouped (Reading, Fiction Writing, Non-Fiction Writing, and Spelling, Punctuation and Grammar (SPaG).

The code-based approach, focusing on English-specific skills, was designed to provide personalised and focused feedback to support pupil learning and progress. A further aim of the intervention was to reduce teachers' marking workload while improving the quality and timeliness of feedback that they were able to provide for students

The intervention comprises the following key elements:

- Training and ongoing support for English teachers

Figure 1. FLASH marking codes.

- Removal of number/letter-based grades from day-to-day written feedback
- Use of FLASH codes in classroom teaching and learning activities
- Embedding FLASH in lesson planning, e.g. in the development and use of schemes of work and teaching resources
- Promoting and supporting development of pupils' metacognitive skills (e.g. planning, self-evaluation).

Two English teachers from each school (including the subject leader for English) were selected to attend three training sessions run by the development team. The first of these took place before the start of the academic year (in July 2018) and the staff involved were then responsible for cascading the training to all department staff who would be working with Year 10 pupils in the following academic year. All staff with GCSE English classes were expected to use the intervention with their pupils. This included sharing FLASH codes as part of lesson/learning objectives, using FLASH Marking when providing feedback on pupils' written work, and tracking pupil progress using the codes. To support the cascading, training materials and resources were provided to be used with staff in schools. All intervention schools also received adaptable FLASH code sheets in the first training session. The training sessions lasted for a day each and were held in June/July 2018 (prior to the start of the trial), September/October 2018 (during the initial months of the trial), and in July 2019 (at the end of the first year of the trial).

The FLASH Marking approach allows schools to develop and adapt the codes according to the marking criteria of the exam boards they were working with, the literary texts and topics being studied, and the existing marking/feedback policies in schools. Examples of how the code sheets could be used were shared by the developers during the training sessions. The developers also shared 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. The aim of these plans was to encourage teachers to embed the intervention into their practice from the outset, and to consider which FLASH codes (skills) could be addressed best through different topics. These resources provided stimulation for discussions at the training sessions, and ideas for planning when teachers were back in school. In the latter two sessions, the developers also shared examples from intervention schools about how

the codes were being used by different teachers. These provided 'real life' models of good practice for other teachers to adopt or use.

A range of resources were developed to facilitate the introduction of FLASH Marking. The code sheets (Figure 1) were provided to all intervention schools at the start of the trial. Some schools opted to adapt their codes (as noted above) or only used a sub-set of codes deemed appropriate for their pupils. Examples of short- and medium-term planning documents and examples of lesson resources were also collated and developed the learning objectives or activities, and a scheme of work, which covered all of the Year 10 curriculum topics and their corresponding FLASH codes. The purpose was to encourage teachers to embed the FLASH codes across their learning objectives, activities and longer-term mapping and planning in Key Stage 4, and to consider which FLASH codes (or skills) could be addressed best through different topics.

Similarly, the bespoke support provided by the development team was accessed and used by schools in different ways. This permitted flexibility and variation is relatively unusual when evaluating 'packaged' interventions such as this. A number of core requirements remained in place in order to assess fidelity to the programme but the developers felt that the more tailored approach could lead to increased engagement and motivation with FLASH. This is a test of a 'template' intervention.

Intervention schools received regular support from the development team throughout the trial via telephone and emails. The development team also visited schools to observe FLASH being used and to discuss and support intervention implementation with subject leaders and teachers. New teaching staff in intervention schools were offered training in FLASH when joining. Teachers were also given access to the web portal, 'Trello', where they could share videos, models of assessed work, and curriculum resources. This was introduced with a view to developing a community of practice around FLASH Marking. As part of their approach to ensuring fidelity, the development team planned to visit intervention schools at least once during the trial, but due to the Covid-19 pandemic and other reasons (staffing, inspection visits) only 32 out of 52 intervention schools received visits.

Methods

This was a two-arm randomised control trial (RCT) with randomisation at the school level. The research question being addressed in this article is:

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

The sample

The sample included 103 schools (with 990 English teachers) and their Year 10 pupils. Of these schools, 52 were randomised to the intervention group, and 51 to 'business-as-usual' (continuing with their usual assessment and feedback practices). The intervention and control schools were similar in characteristics in terms of school type, Ofsted

inspection ratings, location type, size, proportion of FSM-eligible (an indicator of low family income) and EAL (English as an Additional Language). Pupils in the two groups were also balanced in terms of prior Key Stage 2 attainment (end of primary school assessment).

Over the course of the two years, six intervention schools withdrew from the trial, mainly due to staffing changes, both at department and senior leadership level. These schools were not sent the second teacher workload survey. No control group schools withdrew from the trial.

The timing of the trial is significant. The latter stages of this study took place in spring-summer 2020, during the first year of the Covid-19 pandemic. The school closures and subsequent cancellation of externally assessed exams had a major impact on the curriculum and pedagogy for Key Stage 4 pupils, and the lives of teachers too. We acknowledge this as a consideration throughout our analyses and comment further on this in the project limitations section below.

Impact outcomes

The impact of FLASH Marking on teacher workload was estimated using a bespoke pre- and post- online survey developed by the evaluation team. The pre-intervention survey included a total of 20 items adapted from the Department for Education instrument used in the Workload Challenge project (DfE, 2016). The survey asked teachers about the number of hours that they spent on different teaching duties (e.g. feedback and marking, planning lessons, or communicating with parents) and the total hours spent on teaching and learning activities in their most recent full working week. Participants 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 post-survey was administered during the spring term of the second year of the intervention (16–18 months after beginning the trial). It had the same items as the pre-intervention instrument, but also included two additional questions about experiences of using FLASH Marking for teachers in intervention schools. The workload questionnaire can be viewed in the technical appendices to the project report (Morris et al., 2022).

Analyses

The impact of the intervention on teacher workload (number of hours spent on various activities) was expressed in terms of 'effect' sizes computed as the difference between the intervention and control groups, divided by the overall standard deviation. We then looked at the gain (or loss) in working hours from the first survey to the second. For categorical variables, such as teachers' attitudes to their workload, we compared the relative odds of the likelihood of the intervention group and control group reporting an outcome. These are expressed as odds ratios.

There were some missing cases in the second survey, partly due to Covid-19 and the associated lockdown, and staff turnover. This loss of data is clearly not random, therefore approaches such as weighting and multiple imputation are not used here. Instead, we reported the level of missing data, and compared the pre-test scores of those missing the post-test to see if they differed between the two groups (Gorard, 2020).

Because not all of the teachers who completed the second survey had completed the first survey, we ran two analyses. The primary analysis is based only on those teachers who responded to both surveys ($n = 218$). To match responses from the first survey to the second survey, teachers were asked for their names. However, only around 50% provided a readable name and, with the low response to the second survey, this means that we were only able to match 218 responses (102 intervention and 116 control). The main analyses are therefore based on these cases as they provide the fairest (internally valid) comparison between the groups.

The second analysis involves all cases with responses to either pre- or post-intervention surveys. Of the initial 990 teachers, 833 completed the first survey (415 in the intervention group, and 418 in the control schools), a response rate of 84%. In the second survey only 358 teachers (159 in the intervention group, and 199 in the control schools) responded. We believe that the lower response from the second survey was primarily due to the outbreak of the Covid-19 pandemic and the ensuing lockdown; however, there are also other possible reasons, including intervention/questionnaire fatigue and high teacher workload.

Process evaluation

In addition to the data collected for the impact evaluation, we also carried out a detailed and in-depth process evaluation. This involved detailed exploration of feedback and marking in 16 case study schools (13 intervention and three control), along with data collated from items on the teacher questionnaires, a short student survey and observations of training sessions. The case study work involved observing Key Stage 4 English lessons and conducting interviews with pupils and staff involved in the trial. Two visits were made to each case study school per year of the trial (with the exception of some schools who had visits scheduled after the Covid-19 lockdowns). The aim of the visits to intervention schools was to learn, in more detail, about how the intervention was being implemented, the extent to which delivery complied with training, and any challenges faced by schools in using FLASH. We also collected feedback from staff and students about their experiences and attitudes relating to the intervention. As a comparison, we also visited three control schools with a view to understanding the practices in these schools, and the extent to which these were similar/different to FLASH Marking. We also received feedback on FLASH from 474 pupils via a questionnaire survey administered in eight intervention schools, and via the teacher workload survey which included a small number of items about teachers' experiences and attitudes towards FLASH.

Data from the observations, interviews/focus groups and survey were analysed thematically, and numerical data from the staff and pupil questionnaires was analysed using simple, descriptive statistics.

Results

Impact on teachers' working hours

As not all teachers completed both surveys, we first compared the results for the 218 teachers whose surveys could be matched, and then all teachers who completed either of

Table 1. Number of hours worked in the last week, matched respondents ($n = 218$).

	Mean hours total pre-intervention	Standard deviation	Mean hours total post- intervention	Standard deviation	Gain	Standard deviation	'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	-

Note: FLASH teachers also reported greater reduction in hours spent marking compared to the control group ('effect' size = −0.17).

Table 2. Number of hours worked on pupil marking and feedback in the last week, matched respondents.

	Mean hours marking pre-intervention	Standard deviation	Mean hours marking post- intervention	Standard deviation	Gain	Standard deviation	'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	-

the surveys. In general, FLASH Marking appears to have a positive impact in reducing teachers' workload. Tables 1 and 2 show that the two groups were reasonably well balanced at the outset in terms of total hours worked and hours spent marking (as reported by the teachers). Both groups reported a decline in workload, perhaps because of Covid-19, but FLASH teachers reported a bigger reduction in their total working hours compared to the control teachers after the intervention ('effect' size = −0.16).

We cannot be certain of the reason that control school teachers also experienced a reduction in workload and hours spent marking. We did find evidence, however, that during the course of the trial some of these schools were also implementing strategies to reduce workload; it would appear that these approaches were not similar to FLASH Marking and did not have the same level of impact (as discussed further below).

Including all participants who completed either of the two surveys or both, the analysis shows that before the intervention FLASH teachers worked slightly fewer hours in total per week than those in the control group (Table 3). This could be due to different contracts and job status (such as leadership roles) but could also be due to sampling variation. However, after the intervention, FLASH teachers reported working even fewer hours compared to the control ('effect size'= −0.25), suggesting a greater reduction in hours worked than control teachers. It must be noted that these responses are not all from the same cases.

For comparison, those staff who responded only to the first survey reported mean working hours of 44.72 hours (standard deviation 15.62), with a mean of 45.33 for the intervention group and 44.13 for the control. This suggests that those missing the post-test in the two groups did not differ substantially from the overall mean.

Table 3. Number of hours worked in the last week, all respondents.

	Mean hours total pre-intervention	Standard deviation	'Effect' size	Mean hours total post-intervention	Standard deviation	'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	–

Table 4. Number of hours worked on pupil marking and feedback in the last week, all respondents.

	Mean hours marking pre-intervention	Standard deviation	'Effect' size	Mean hours marking post-intervention	Standard deviation	'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	-

Similarly, considering only the hours spent on marking and pupil feedback (Table 4), the groups were slightly unbalanced at the outset, but in the second survey FLASH teachers reported substantially less marking than the control ('effect' size −0.27).

For comparison, teachers who responded only to the first survey (i.e. those missing post-survey) reported a mean of 7.44 hours spent marking (standard deviation 4.81). This is slightly lower than the average for all respondents (7.86 hours), but the difference is not large. Due to missing data in the post-test, caution is needed in interpreting these findings, but taken together these four tables suggest that FLASH Marking had a small positive impact in reducing teachers' workload in terms of time spent on teaching, marking, and giving feedback.

Extensive feedback from teachers (via the workload survey, case study visits and interviews) supports these findings. Several Heads of Department commented on the impact of FLASH Marking in the 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.

> I think it has taken a while for every teacher to fully get on board. Early adopters are ahead in their journey therefore. 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.

Attitudes to workload and FLASH marking

To understand whether FLASH Marking had made a difference to teachers' attitudes towards workload, we compared the responses of 358 teachers who completed the second survey. On average, FLASH teachers were twice as likely to report that their overall workload was more acceptable than the control group (Table 5). If we combine both levels of agreement, then the odds ratio is $(19 \times 78)/(12 \times 61)$ or 2.02 in favour of the intervention group. We also see that a fifth of the intervention group chose the middle option (neither agree nor disagree) compared with just 10 per cent of the control group.

Table 5. Percentage of respondents in each group agreeing that they have an acceptable workload.

	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 6. Percentage of respondents in each group responding to statements about the amount of marking they do.

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

FLASH teachers were 1.62 times more likely than the control teachers to report that their marking workload was acceptable (Table 6), aligning with survey responses where 76% of teachers believed that FLASH had reduced their workload. While these are potentially positive findings, we can also see in Table 6 that over a third (37%) of intervention teachers still reported that their marking workload was too high, indicating that FLASH Marking did not have this desired impact across all schools or teachers.

Teachers' feedback on the training, intervention, and support of staff

To yield the most benefit from FLASH Marking, it is important that teachers are trained to deliver it effectively. Our process evaluation showed that the training was well received by teachers. FLASH teachers were generally positive about the approach and the training they received. They tended to agree that the training they received was helpful (77% agreed), and that they would recommend the intervention to other schools (79%). Over three-quarters (77%) thought that the intervention had benefitted pupils, and most Heads of English (91%) were satisfied with their teachers' commitment to FLASH Marking.

Teachers generally reported that the training was high-quality, well-prepared, and informative. Many especially valued the way that the content was underpinned by research evidence and practice. Attendance at the three training sessions was high, with 100% attendance at the first session, dropping to 89% by the third session. All participating schools reported that training was cascaded to the rest of the staff in the school, although how this was delivered differed between schools. One Head of English commented that they would cascade the training of FLASH to teaching assistants (TAs) so that they were more familiar with the approaches used and could better support students in the class.

Teachers were appreciative of the monitoring and regular support offered by the development team. Such support involved visiting schools and sharing resources via the online Trello platform. As an example, one Head of English commented on the monitoring support they received:

> [The FLASH 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 . . . and stayed for our department . . . The feedback that we received was excellent, and to get external verification of our implementation was really reassuring as this process was obviously new to us.

The use of FLASH in the classroom

As part of the process evaluation, we also looked at how FLASH was implemented in practice. All teachers in intervention schools reported a reduction in the use of

alphanumeric grades for formative work, but most still retained some grades in more formal assessments such as mock exams. This was evidenced in the pupils' exercise books, and discussions with pupils. Teachers acknowledged that removing grades altogether was often not easy or feasible in their current contexts. Sometimes they were required by school policy to provide regular grades/marks to pupils, especially for those preparing for their GCSEs. Teachers explained that some parents also expected to see grades regularly, to understand the progress that their children were making.

Across the intervention schools, we observed widespread use of the FLASH Marking codes in classroom-based activities and saw a range of innovative ways that teachers integrated the codes into their lessons. Teachers were given considerable freedom to incorporate FLASH codes into their teaching as they deemed appropriate. They were shown examples by the development team of how this could be done but were also encouraged to think of creative ways to use the codes. We saw regular examples of codes being used by teachers to highlight lesson objectives, to provide success criteria or as feedback to pupils. The codes were also used frequently by pupils, usually for formative assessment purposes (e.g. via self- or peer-assessment). In some schools, the number of codes was reduced to align them more with the exam specification that the school was working with, as illustrated by a comment from a teacher:

> 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.

The use of FLASH codes was often accompanied with colour-coding in written texts where students were 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. These activities were designed to facilitate consistent and accurate code use, and to encourage students 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 students' 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.

Pupils spoke enthusiastically about the use of coloured highlighters to identify the different features of their writing and to spot 'gaps' or areas for development. Pupils also liked the code sheets, which they pasted onto their exercise books. They found them a helpful reminder of what the codes stood for. In some schools these code sheets were on display walls in the classroom or corridors. While many teachers felt that there was value in the colour-coding, some noted that it could be a distraction for some children.

The frequency and quantity of FLASH Marking codes varied both within and across schools. In some schools FLASH was used in every lesson and was apparent in pupils' exercise books. In other schools, FLASH was less visible. Our visits to cases study schools also indicated that more experienced teachers were sometimes reluctant to embed it consistently into their regular teaching practice.

There were some divergent views about whether FLASH is appropriate for lower ability students. A small number of teachers felt that FLASH Marking was not particularly useful for pupils with lower prior attainment or Special Educational Needs (SEN) because the number of codes and the need for the pupils to understand a new system of marking and feedback can be confusing. However, in the majority of schools, we observed examples of FLASH being used as intended with lower attainers in English, and many teachers were very enthusiastic about the outcomes they saw for these pupils:

> No doubt about it, FLASH has saved my bottom set Y11 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.

Use of FLASH in curriculum and planning

Our observations and discussions with English staff indicated that FLASH was being added into schemes of work 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, several teachers told us that they could see value in this kind of longer-term planning. Teachers in two other schools told us that they would like to embed FLASH more fully into their scheme of work documents but were unable to do so due to time and workload constraints. In other schools, teachers adapted their teaching resources (such as PowerPoints and worksheets) to include reference to FLASH Marking codes. In many schools, we saw teachers using a common set of resources so that there was consistency in the way the codes were shared and used across the English department.

Challenges in the use of FLASH Marking

Despite the positive feedback from many intervention schools, some schools reported that adopting FLASH Marking was challenging at the outset. Encouraging teachers and pupils to change their traditional marking/feedback habits was found to be initially difficult in some settings. One Head of English reflected that it had been 'a bit of a challenge to wean our team off [graded] marking'. Another Head of English explained that:

> 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.

Staff turnover was another challenge and meant that new English teaching staff had to be trained and brought up to speed with the programme. In a two-year trial, it was expected

that a number of teaching colleagues would change roles or leave the school, although some trial schools experienced particularly high rates of turnover and were sometimes reliant on supply staff to cover teacher absence. Heads of Departments were expected to re-deliver the cascade training to new teachers and most did this where possible. Given time constraints and capacity issues, however, they noted that this was not always as detailed as they would have liked and that they did not always have enough time to monitor new teachers' use of the approach.

Support from senior leadership was also seen as important in successful implementation of the programme. Where senior leaders were also members of the English department, we saw increased 'buy in' with FLASH Marking. Where support from senior leadership was absent, teachers described being 'left alone' to roll out and implement FLASH Marking independently. Several schools withdrew from the trial in the second year when they had a change of head teacher; some teachers from these schools reported feeling disappointed as they had invested considerable time and effort into the study.

Impact on pupil outcomes

Although the intention of the trial was also to evaluate the impact on pupils' learning, this was not possible due to Covid-19 lockdown and the cancellation of externally assessed GCSE exams in summer 2020. The project funders determined that teacher-assessed GCSE grades would not be robust enough for measuring attainment as part of the trial, due to their inconsistency and potential for grade inflation. As a result, these analyses were not possible. Instead, we only have teachers' and pupils' reports on the perceived impact on learning.

In the second survey, 77 per cent of teachers felt that FLASH Marking had had positive benefits for students. One Head of Department explained that the FLASH codes had helped pupils to recognise what skills and knowledge were required in the exams. In this way, pupils reportedly understood 'gaps' in their knowledge or skills, and could identify what was needed to gain marks and to improve on their previous performance. Other teachers reported that pupils' outcomes were improving because of FLASH:

> . . . staff are more likely to adapt and plan lessons based on in-class assessment and so lessons are more tailored to students' needs.

> My Y11s (lower ability) have just completed their mock exam and have done so well – they were using their codes to plan and structure their work and have all commented on how helpful it has been.

However, not all teachers were so positive. In some schools, teachers were apprehensive about the academic benefits of the intervention. They were concerned that FLASH Marking had not been implemented properly, and some had combined FLASH with more written feedback or marks as they did not feel confident relying on FLASH Marking alone. One teacher said:

> 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 as more of a chore than a useful tool. Many pupils fail to see how the codes effectively correspond to their work. The

Table 7. 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

> 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.

Some teachers and Heads of Departments provided reflections on the potential value of FLASH Marking to support and develop metacognitive skills. While the developers intended this as an outcome, this was not tested or examined as part of the trial, and it is this not discussed further here.

Pupils' views on FLASH Marking

Pupils themselves reported generally favourable views about the intervention and its influence on their learning (see Table 7)

Most pupils we spoke to had positive views about the approach. They talked about the value of using feedback to improve their work. The general opinion is that FLASH offered a clear, focused approach to improvement and development, and encouraged them to think and reflect on their work. The following comments from the survey were indicative of these views:

> 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 feedback we get afterwards.

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

However, not all pupils found FLASH useful. A small number of higher-attaining pupils, for example, did not find FLASH Marking helpful, and often compared it with previously used approaches. They preferred detailed comments and/or marks/grades and wanted more positive affirmation of the achievements they had demonstrated. One pupil found the approach mechanistic and suggested that it stifled her creativity.

> 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.

What do control schools do differently?

Through our visits to a small sample of control schools (*n* = 3) and our questionnaires with control school teachers, we found that they did not appear to be consistently using approaches that were similar to FLASH Marking. While the marking policies in some

control schools were also aimed at reducing workload, the strategies that we saw did not encompass the multiple elements observed with FLASH. Control schools were not able to access the training or resources associated with FLASH, and from our observations most either continued with their pre-trial practices or undertook some assessment policy amendments to reduce workload. One teacher, who had moved from an intervention school to a control school during the course of the trial, made a very useful comparison of the approach used in the control school with FLASH Marking.

> I have had the benefit of moving schools – both of which have been involved within this trial . . . The Flash Marking system just gave a degree of rigour and formality to what I was using. I then moved to [Control School] . . . I did try to control my 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 exemplar 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.

Of course, other schools might have adopted procedures to reduce the marking workload that were effective. But we saw no pattern, and even if this was the case, the impact would simply have been to reduce the apparent effect size for FLASH.

Possible limitations of the study

Given that this evaluation was disrupted by the Covid-19 pandemic, it is important to outline possible resulting limitations. We note these below, along with methodological challenges which should be considered when interpreting the findings from this study. In the last year of the trial, many schools closed to the majority of pupils due to Covid-19. Despite this, some continued to use FLASH Marking where possible. Teachers had to re-prioritise their work and support pupils through distance learning, often resulting in changes to the curriculum and reduced engagement from pupils. These were not the conditions that FLASH was designed for. Implementation of the intervention in these latter stages was therefore patchy and difficult to monitor. The cancellation of externally assessed national examinations was also a result of the pandemic and meant that we were unable to carry out analysis of pupils' attainment outcomes. Instead, we have relied upon teachers' and students' reports of perceived impact. While helpful, these are not as robust or meaningful for understanding the actual effectiveness of an intervention on attainment.

School closures made it difficult to reach teachers to get them to complete the second survey, and the response rate was, therefore, low. Only 36 per cent of the initial 990 teachers completed the second survey. The main impact analysis on workload was therefore based on only 218 teachers (22%) who completed both pre- and post-surveys. This limits our ability to make more general statements about the impact of FLASH.

Another issue leading to caution is the reliability of teacher responses. As teachers in the intervention arm were not (and could not be) blind to the intervention, knowledge of being in the intervention might have encouraged teachers to respond in a way that they would not otherwise. Our in-depth, longitudinal data from the case study schools was mostly collected from schools and teachers that were, to some extent, engaged with the trial and keen to talk to us. There were some exceptions to this (including one case study intervention school which really struggled to embed FLASH across the period of the trial),

but the selection and involvement of these schools could have implications for the findings as these teachers were perhaps more likely to report a positive view of the intervention than might be found for schools as a whole.

Conclusion

This study was the first evaluation of the FLASH Marking intervention. Our findings suggest that this approach is potentially effective in addressing teachers' workload, with intervention teachers reporting a reduction in the number of working hours per week compared with teachers working in the control schools (a difference of four hours), and a reduction in the time spent on marking and feedback (a difference of 1.5 hours). Teachers 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. In terms of pupils' learning, teachers noted improved progress and engagement with feedback, and the pupils who responded to our survey were generally positive about the potential impact that FLASH had on their skills and knowledge in English.

FLASH Marking is somewhat unusual in the potential for the intervention to be modified and adapted for specific school contexts, pupil needs or exam board requirements. This provides a contrast to more rigid approaches which require closer adherence to a specified set of behaviours or activities. Teachers appreciated this flexibility and our evaluation data suggests that it supported engagement and motivation to use FLASH. This bespoke nature of FLASH Marking does make measuring fidelity and compliance more complex in the context of an RCT; nevertheless, we believe that there is scope for further exploration of these more 'flexible' approaches to improvement, and for combining a range of methods (including trials) for assessing their implementation and impact.

Due to the lack of externally assessed GCSE results at the time of the trial, the impact evaluation of pupil attainment outcomes was not conducted. Future research should consider using standardised tests to provide a robust evaluation of the impact of the FLASH approach (or similar code-based methods) to marking and assessment on pupils' attainment in English. In line with recent reviews of feedback in schools (Elliott et al., 2016; Newman et al., 2021), we also support the view that further research is needed more generally on the effects of written marking, particularly those approaches which have the potential to support teachers through high-quality professional development and workload reduction.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Rebecca Morris is an Associate Professor in the Department of Education Studies, University of Warwick. Her research interests include education and social policy, the teacher workforce and teacher education, and English and literacy.

Stephen Gorard is the Director of the Evidence Centre for Education, Durham University, and Fellow of the Academy of Social Sciences. His work concerns the robust evaluation of education, focused on issues of equity and improvement. He is author of 30 books and over 1,000 publications, making him the most published/cited UK education author in the Web of Science over the past 50 years. He is currently funded by the ESRC to examine teacher supply, and to investigate the impact of ethnic diversity of the teaching workforce, and by DfE to evaluate the impact of Glasses in Classes in Opportunity Areas.

Beng Huat See is Professor of Education Research at the Durham University School of Education. She is a nominated academician of the Academy of Social Sciences. Her current research interests focus on improving teacher supply, teacher development and teacher quality as well as identifying ways to enhance students' learning and wider outcomes. Her expertise is in the use of multiple designs: rigorous reviews and synthesis of evidence in education, evaluation of large-scale RCTs and re-analysis of secondary data.

Nadia Siddiqui has academic expertise in education research and equity in education. She has led important education research projects contributing evidence for education policy in England, and education evaluation studies of widely used catch-up learning programmes in primary schools. Her research interests lie in exploring the stubborn patterns of poverty and inequalities through population data sets, large scale surveys and cohort study datasets. By using these secondary data resources she investigates the indicators of disadvantage that determine children's academic attainment, wellbeing and happiness, and access to pathways for successful life.

ORCID

Rebecca Morris 🆔 http://orcid.org/0000-0002-1699-4172
Stephen Gorard 🆔 http://orcid.org/0000-0002-9381-5991
Beng Huat See 🆔 http://orcid.org/0000-0001-7500-379X
Nadia Siddiqui 🆔 http://orcid.org/0000-0003-4381-033X

References

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Churches, R. (2020). *Supporting teachers through the school workload reduction toolkit*. 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

DfE. (2016). *Teacher workload survey 2016*. https://www.gov.uk/government/publications/teacher-workload-survey-2016

Elliott, V., Baird, J., Hopfenbeck, T., Ingram, J., Thompson, I., Usher, N., Zantout, M., Richardson, J., & Coleman, R. (2016). A marked improvement? *A review of the evidence on written marking*. EEF, Written marking | EEF. educationendowmentfoundation.org.uk

Elliott, V., Randhawa, A., Ingram, J., Nelson-Addy, L., Griffin, C., & Baird, J. (2020). *Feedback in action: A review of practice in English schools*. EEF. https://d2tic4wvo1iusb.cloudfront.net/documents/guidance/EEF_Feedback_Practice_Review.pdf

Gorard, S. (2020). Handling missing data in numeric analyses. *International Journal of Social Research Methods*, *23*(6), 651–660. https://doi.org/10.1080/13645579.2020.1729974

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Independent Workload Review Group. (2016). *Eliminating unnecessary workload around marking*. DfE. https://www.gov.uk/government/publications/reducing-teacher-workload-marking-policy-review-group-report

Kime, S. (2018). *Reducing teacher workload: The 'rebalancing feedback' trial*. https://dera.ioe.ac.uk/31210/1/Cheshire_Vale_-_Reducing_teacher_workload.pdf

Klapp, A. (2015). Does grading affect educational attainment? A longitudinal study. *Assessment in Education: Principles, Policy & Practice*, *22*(3), 302–323. https://doi.org/10.1080/0969594X.2014.988121

Koenka, A. C., Linnenbrink-Garcia, L., Moshontz, H., Atkinson, K. M., Sanchez, C. E., & Cooper, H. (2021). A meta-analysis on the impact of grades and comments on academic motivation and achievement: A case for written feedback. *Educational Psychology*, *41*(7), 922–947. https://doi.org/10.1080/01443410.2019.1659939

Lipnevich, A., & Smith, J. (2009). Effects of differential feedback on students' examination performance. *Journal of Experimental Psychology*, *15*(4), 319–333. https://doi.org/10.1037/a0017841

McDonald, R. (2021). *What does effective whole-class feedback look like?* Impact – Journal of the Chartered College of Teaching. https://impact.chartered.college/article/effective-whole-class-feedback-english/

Morris, R., Gorard, S., See, B. H., & Siddiqui, N. (2022). *FLASH marking evaluation report*. EEF.

Motteram, G., Choudry, S., Kalambouka, A., Hutcheson, G., & Barton, A. (2016). *ReflectED. Evaluation report and executive summary*. Education Endowment Foundation.

Muijs, D., & Bokhove, C. (2020). Metacognition and self-regulation. *Evidence Review*. EEF. https://educationendowmentfoundation.org.uk/public/files/Metacognition_and_self-regulation_review.pdf

Newman, M., Kwan, I., Schucan Bird, K., & Hoo, H.-T. (2021). *The impact of feedback on student attainment: A systematic review*. EEF. https://d2tic4wvo1iusb.cloudfront.net/documents/guidance/Systematic-Review-of-Feedback-EPPI-2021.pdf

Perryman, J., & 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. https://doi.org/10.1080/00071005.2019.1589417

Riley, K. (2020). *Whole class feedback: Reducing workload, amplifying impact and making long-term change in the learners*. https://myhodandheart.wordpress.com/2020/10/09/whole-class-feedback-reducing-workload-amplifying-impact-and-making-long-term-change-in-the-learners

Speckesser, S., Runge, J., Foliano, F., Bursnall, M., Hudson-Sharp, N., Rolfe, H., & Anders, J. (2018). *Embedding formative assessment: Evaluation report and executive summary*. https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/embedding-formative-assessment

Toropova, A., Myrberg, E., & Johansson, S. (2021). Teacher job satisfaction: The importance of school working conditions and teacher characteristics. *Educational Review*, *73*(1), 71–97. https://doi.org/10.1080/00131911.2019.1705247

Zhang, B., & Misiak, J. (2015). Evaluating three grading methods in middle school science classrooms. *Journal of Baltic Science Education*, *14*(2), 207. https://doi.org/10.33225/jbse/15.14.207