Digital Feedback in Primary Maths

Evaluation Report

September 2019

Alex Sutherland, Miriam Broeks, Megan Sim, Elena Brown, Eleftheria Iakovidou, Sonia Ilie, Hannes Jarke, Julie Belanger

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 (formerly Impetus Trust) and received a founding £125m grant from the Department for Education.

Together, the EEF and Sutton Trust are the government-designated What Works Centre for improving education outcomes for school-aged children.

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

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

0207 802 1653

jonathan.kay@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Contents

About the evaluator

The project was independently evaluated by a team from RAND: Alex Sutherland, Miriam Broeks, Megan Sim, Elena Brown, Eleftheria Iakovidou, Sonia Ilie, Hannes Jarke, Julie Belanger.

The lead evaluator was Alex Sutherland.

Contact details:

Name: Miriam Broeks
Address: RAND Europe, Westbrook Centre, Milton Road, Cambridge, CB4 1YG, United Kingdom
Tel: +44 1223 353 329
Email: mbroeks@rand.org

Executive summary

The project

The Digital Feedback in Primary Maths programme aimed to improve the feedback provided by primary school teachers using a tablet application called Explain Everything, diagnostic assessments, and training on effective feedback. In this trial, it was expected to increase the maths attainment of pupils in Years 4 and 5. The programme was developed and delivered by KYRA Teaching School Alliance.

Instead of providing written comments, teachers used the app to record video of themselves providing verbal feedback on pupils' work or record themselves annotating photographs of pupils' work. Pupils were able to access and replay the feedback videos using tablets available in class. Teachers could address common misconceptions by making recorded feedback available to all pupils in the class. The delivery team also provided teachers with diagnostic assessments to assess pupils' understanding of a topic. These assessments were very short, consisting of only three or four questions. It was suggested that teachers use both the digital feedback and diagnostic assessments two or three times a week.

The delivery team supported teachers to adopt these approaches by providing initial training, setting up Professional Learning Communities (PLCs), and arranging coaching. Participating teachers attended a one-day training session on the principles of effective feedback and on using the app. Teachers were assigned a PLC of 8 to 12 teachers from different schools who were expected to meet once a month to discuss the implementation of the intervention. One teacher in each participating school was designated as a Research Lead, responsible for providing coaching to their colleagues using the programme. The project team provided Research Leads with a half-day training session on their coaching role. It was intended that Research Leads would also meet once per term to receive additional training.

The impact of Digital Feedback on maths was estimated using a randomised controlled trial involving 34 schools, 108 classes, and 2,564 pupils. Classes were randomised to either receive the intervention or to a control group which carried on with business-as-usual teaching. Maths attainment was measured using ACER's Essential Learning Metric and the level of disaffection with school was measured as a secondary outcome. A process evaluation used surveys, interviews, observations of training and PLC sessions, and analysis of feedback videos, coaching records and diagnostic assessments to examine how the programme was implemented. The delivery of the programme started in September 2017 and concluded in June 2018.

Key conclusions	
1.	There is no evidence that Digital Feedback in Primary Maths had an impact on pupils' maths outcomes. This result has a moderate to low security rating.
2.	There is no evidence that the programme has an impact on the maths outcomes of pupils eligible for free school meals.
3.	The programme, as designed, was not easily implemented by teachers and schools. There was evidence that teachers and Research Leads sometimes did not attend training, have coaching sessions, or provide regular digital feedback in the classroom.
4.	Both the diagnostic assessment and training in effective feedback were reported as more useful by teachers than the use of the app to provide digital feedback. Evidence from surveys, PLC observations, and an interview indicated that the diagnostic assessments were eventually being used as a global classroom assessment tool.
5.	There were significant challenges with identifying which pupils were in the treatment group and which were in the control. This meant that 431 pupils could not be included in the main analysis. This presented a small threat to the security of the trial.

EEF security rating

These findings have a moderate to low security rating. This trial was an efficacy trial, which tested whether the intervention worked under developer-led conditions. It was difficult to ascertain whether some pupils were in the treatment or control group. This meant that the evaluator was only able to analyse the data of 83% of the pupils for the primary outcome. There were various other issues that limited the security of the trial, including a high risk of

contamination between control and intervention participants. This was due to the randomisation of classes rather than schools and some teachers in the control group being given tasks related to the intervention.

Additional findings

There was some evidence that pupils receiving the programme were less disengaged with school than pupils in the control group. However, a large number of pupils did not complete this outcome measure which seriously weakens the security of this finding.

Teachers and Research Leads initially reported that the training prepared them well to deliver the intervention. However, half way through the intervention delivery most survey respondents reported feeling only somewhat prepared (rather than very prepared) to implement Digital Feedback, with some Research Leads reporting not feeling prepared at all. Similarly, PLC meetings were overall found to be 'somewhat useful' or 'very useful', but attendance at PLC meetings decreased over time. Coaching meetings appear to have happened on an irregular basis, with some teachers reporting not having met with their Research Lead at all.

The quality of teacher's implementation of the intervention in the classroom was mixed. The majority of teachers reported using diagnostic assessments to focus teaching on areas that the class seemed to be struggling with and to track progress made by the class as a whole over the year. However, survey respondents noted that embedding the use of Digital Feedback in daily practice was difficult. Intervention survey responses showed a general decrease in the amount of digital feedback given by teachers as the trial progressed. Potential factors hindering the provision of digital feedback identified by participants were lack of equipment or the cost of equipment, difficulty in using—and cost associated with—the Digital Feedback application, and the time taken to provide video feedback.

Future versions of the approach could revise the programme to focus on the aspects that were feasibly implemented by teachers. This could involve exploring whether the training on its own would be effective without the digital element.

Cost

The cost of Digital Feedback was very low. If delivered over three years, Digital Feedback would cost an average of £14.51 per pupil per year.

Impact

Table 1: Summary of impact on primary outcome

Outcome/ Group	Effect size (95% confidence interval)	Estimated months' progress	No. of pupils	P value	EEF security rating	EEF cost rating
Mathematics	-0.04 (-0.17, 0.09)	0	2,133	0.507	🔒🔒🔓🔓🔓	£ £ £ £ £
Mathematics (FSM pupils)	0.04 (-0.24, 0.32)	0	635	0.061	n/a	£ £ £ £ £

Introduction

Background evidence

Digital Feedback in Primary Maths is a school-developed approach to improving teachers' diagnosis and feedback skills when teaching maths in primary schools. The approach of using a tablet to facilitate effective feedback was originally developed and tested by James Siddle for the KYRA Teaching School Alliance (Siddle, 2015). The intervention was developed with funding from the Education Development Trust (formerly Centre for British Teachers). The intervention focuses on five elements, namely, training, high quality feedback, using digital feedback, Professional Learning Communities, and coaching.

Research shows that teacher feedback is among the most powerful factors influencing pupil achievement (Hattie, 2008). At the same time, an EEF review found that there is great variability in the effect sizes (ES) reported in meta-analysis studies examining the impact of feedback and related constructs (EEF, 2018a). This indicates that some forms of feedback may be more effective than others at improving student learning. Reports have also suggested that there is a need for more research to better understand which feedback approaches are most effective (EEF, 2016). For example, the EEF's review of marking (EEF, 2016) found some evidence to suggest that frequent verbal feedback might be more effective than written feedback.

Evidence has emerged that suggests computer-assisted feedback or feedback provided in the form of video or audio are potentially among the most effective approaches to delivering feedback in the classroom (Hattie and Timperley, 2007). In fact, initial evidence from a small randomised controlled trial (RCT) undertaken in 11 school classes by Siddle (2015) found positive effects of an app-based video feedback approach, with an overall small effect on all pupils and a moderate effect for disadvantaged pupils, that is, those with special educational needs and disabilities (SEND) or free school meal (FSM) eligibility (for RCT poster please see Appendix F). These results suggested the need for a larger and independent RCT.

It should be noted that while the RCT undertaken by Siddle (2015) focused on writing tasks, the current trial focuses on maths because unpublished results by Siddle (2015) suggest that the impact of digital feedback on mathematics may be greater than on writing. A potential explanation for this is that mathematics lends itself well to the use of diagnostic assessments as there are well-documented common misconceptions in maths which can get in the way of pupils reaching their learning goal (Ketterlin-Geller and Yovanoff, 2009; Tobey, 2017). The diagnostic assessments are developed to detect these common misconceptions and thus allow teachers to provide specific feedback to specific pupils. Moreover, it was believed that assessing maths at scale would be less prone to measurement error than, for example, writing tasks, as suggested by the recommendations made by Siddle (2015) (please see Appendix F).

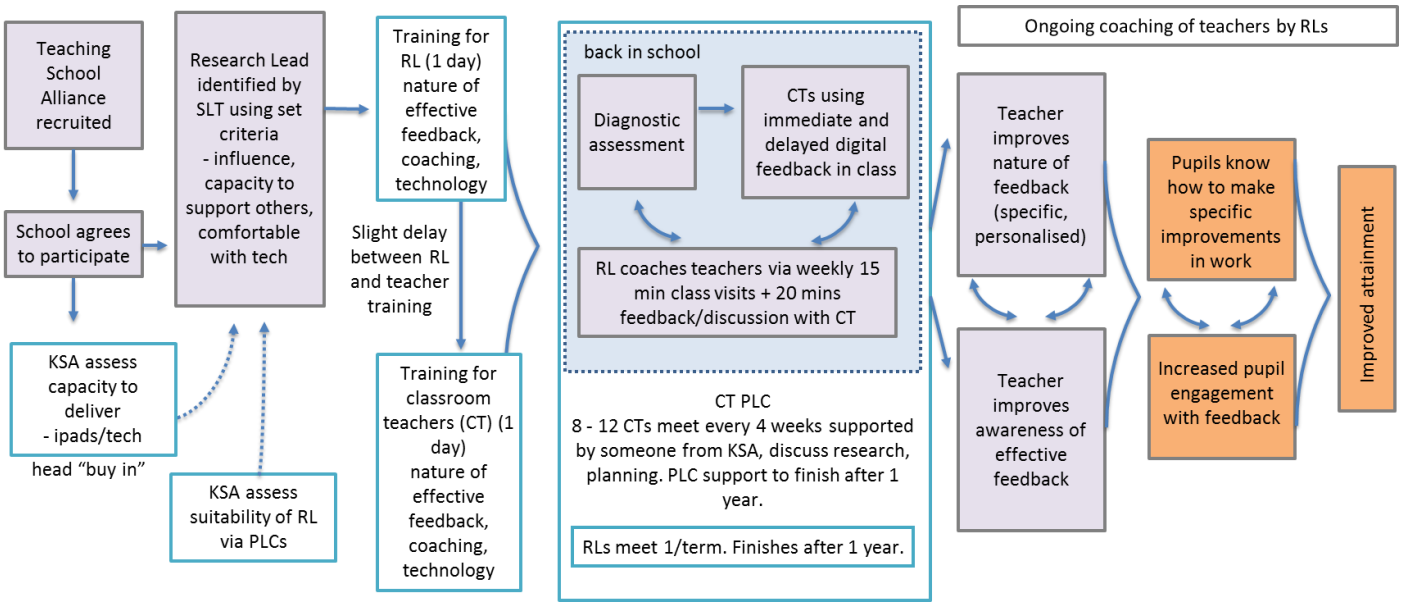
Intervention

Digital Feedback in Primary Maths aims to increase the *timeliness*, *specificity,* and *relevance* of teachers' feedback using a digital app and tablet, making it easier for pupils to review and respond to feedback.

Digital Feedback is a complex intervention that has five elements: training, providing high quality feedback, using digital feedback, Professional Learning Communities (PLCs), and coaching. The *training* seeks to equip teachers with knowledge on how to give *high quality feedback* and on how to provide it in an effective *digital* format. The *PLC meetings* and the *coaching*—which can be considered a form of CPD—aim to reinforce and advance this knowledge. PLC coordinators and Research Leads provide ongoing support on the use of effective digital feedback to teachers in PLC meetings and coaching sessions respectively. Ultimately, the three elements of *training*, *PLC meetings*, and *coaching* aim to improve the quality and effectiveness of feedback that teachers provide. Furthermore, through giving and recording *better* feedback, the intervention intends that pupils can review explanations of an answer at their own pace and have a repository of questions and answers that can serve as a resource for revision. The intervention requires that schools have the IT resources to implement it, classroom teachers that are appropriately trained, and that senior staff are available to coordinate, coach, and take a lead on monitoring its use.

Figure 1 sets out the intervention logic model developed during pre-intervention meetings between the Education Endowment Foundation (EEF), KYRA Teaching School Alliance, and the evaluators (RAND Europe).

Figure 1: Digital Feedback in Primary Maths logic model



The above model presents what was *intended* to be delivered by the intervention team prior to the start of the trial. Following the trial's implementation process evaluation (IPE), data showed that not all these elements were implemented as originally planned. Elements that differed were: the first Research Lead's training session was half a day rather than a full day, the additional Research Leads' training did not take place once per term, and in most schools the meetings between Research Leads and teaches were irregular and did not take place on a weekly basis.

It should also be noted that while the intervention elements and guidelines were the same for all schools, schools were given freedom to implement Digital Feedback how they wished during the school year as detailed below. In what follows, the five intervention elements are described in more detail.

Training

Prior to the start of the intervention, participating teachers and Research Leads in each school received training in the use of the Explain Everything app and in giving effective feedback. Participating teachers attended a one-day training session, while Research Leads attended a half-day training session (though, as discussed, this was initially intended to be a one-day session). The training sought to provide training in the principles of high-quality feedback and familiarise teachers and Research Leads with the app. In addition to this content, the focus of the Research Leads' half-day training was on the coaching role. Teacher training was longer since it included hands-on exercises, for example using the app.

Apart from the first pre-intervention half-day training for Research Leads, the intention was that research leads would meet once per term to receive additional training (there are three terms in the English school year). In practice, there were two additional half-day training sessions available (one in the 2017 Autumn term, the other in the 2018 Spring term). These two sessions focused on the research underpinning effective feedback and the associated pedagogy similar to the focus of the PLCs. Due to low sign-up rates to the Spring term session, the delivery team provided additional individual school visits to support RLs in some schools.

Principles of high-quality feedback

Digital feedback in Primary Maths is based on the concept of effective feedback focusing on the principles set out in a review by Hattie and Timperley (2007). This review describes how feedback can be used to foster metacognition[1] by

[1] The Oxford dictionary defines metacognition as "awareness and understanding of one's own thought processes". Examples of questions a teacher can ask to prompt metacognition are: 'What strategies do you use in solving questions like these?', 'Why does

increasing pupils' understanding of their performance in relation to the learning goal. Feedback can provide the pupil with tools to reach their learning goal by setting out specific steps that the pupil can take to reach the learning goals. Thus, the teacher's role in providing feedback is to support students in monitoring their progress and help them in developing an understanding of the steps and strategies that can be employed to reach a goal.

While the types and frequency of feedback was up to the discretion of the teachers (see section below 'The use of digital feedback' for details), generally the feedback was intended to be of the following nature:

1. **Immediate feedback**—videoed during lessons and given to students straightaway—seeks to help pupils progress with their mathematics work while in class. It may summarise a conversation a pupil and teacher had or may focus on the diagnostic assessment that was taken in class.
2. **Delayed feedback**—videoed outside of lessons, for example, marking homework with a narrative for a specific pupil—focuses on progress pupils make in their mathematics books which follow the curriculum. As the feedback can be reviewed at later points in time, it seeks to help students with the next lesson and is therefore forward looking.

In both cases, the aim of the feedback is to help pupils clarify their own understandings and misunderstandings and help them to monitor their own learning

The use of digital feedback

The use of new technology is innovative but the underlying model is one of improving the feedback that teachers give their pupils (that is, providing effective feedback). It was expected that the use of a digital platform for feedback would allow teachers to improve their feedback in two ways (Siddle, 2015). Firstly, it facilitated the ease of administering diagnostic feedback, which ultimately allowed teachers to provide specific feedback tailored to the needs of individual pupils. Secondly, it allowed teachers to provide the two different types of feedback outlined above. There was also an expected efficiency gain by answering a specific question for pupils, for example, in a group-learning exercise, every pupil in that group then has the feedback available to them and pupils can replay feedback without the need to ask the teacher the same question.

As part of the intervention, teachers were provided with an app ('Explain Everything' app) that allowed them to provide this feedback. The Explain Everything app enables the recording of visuals and voice, either by recording videos or by recording voice while annotating photographs. While this app was provided (free of charge) by the Digital Feedback team for the trial, teachers were free to provide digital feedback how they wished during the school year (that is, using or not using the app to record feedback videos), but all teachers received the same training. During the training, teachers were advised to use digital feedback two to three times a week (but teachers could change this amount depending on the needs of the class, as well as the length[2] of the digital feedback videos) and they were encouraged to annotate and elaborate over a photograph of pupils work or a model when recording feedback. Pupils were able to access the feedback videos using the tablets available in class.

Diagnostic assessments provided by the intervention team were designed to allow teachers to quickly assess their pupils' understanding of a topic, for example, through multiple choice questions (MCQs). These assessments were very short, consisting of only three or four questions to enable teachers to quickly assess pupils' development levels (please see Appendix E: IPE details for an example). These assessments could be used as often as teachers deemed useful and necessary, and helped teachers identify pupils that had a greater learning need and tailor the delayed feedback to the needs of those pupils. The training guidelines suggested using diagnostic assessments two to three times a week.

Teachers were also expected to provide feedback to help those pupils struggling with a topic by using tablets to record video summaries of their feedback, rather than writing down comments (delayed feedback). The tablets were loaded

that answer make sense to you?', 'What things should you pay careful attention to when solving questions like these?' (Crawford, 2018; Teach Thought Staff, 2017).

[2] Teachers were free to define the length of the videos, that is, to make them as long as needed to explain the task. Based on the feedback videos reviewed as part of the IPE, the videos tended to be short—between one and three minutes in most cases.

with the Explain Everything app specifically designed for this purpose. The feedback was given throughout the school year on all types of mathematics classwork.

Professional Learning Communities

Teachers were required to attend monthly Professional Learning Communities (PLCs). The PLCs were group sessions of around 8 to 12 classroom teachers from different schools who met to discuss the implementation of the intervention, moderated by a PLC coordinator from Bishop Grosseteste University employed by the intervention team. The focus of the PLCs was the research underpinning effective feedback and the associated pedagogy.

At the end of the first meeting that teachers attended they created an action plan. The action plans set out targets for the month and were revisited and updated at the end of every PLC meeting. There were eight PLC meetings which were planned to take place every month. The content of the action plans varied depending on the needs of individual teachers and their classes. The action plans were incorporated into the coaching sessions teachers had with the Research Lead in their school.

Coaching

Teachers received coaching from Research Leads based in their school. Research Leads were selected by participating schools with guidance from KYRA. They were generally either experienced teachers or senior staff able to provide peer support to teachers assigned to implement Digital Feedback. Research Leads also received training in using the app and in giving feedback making them qualified to support teachers on an ongoing basis.

Coaching of teachers by Research Leads was intended to consist of two elements. First, Research Leads conducted weekly 15-minute class visits to observe feedback practices and the use of the app. Second, Research Leads and teachers met on a weekly basis to allow Research Leads to provide feedback from their observations and discuss this with the teacher. Research Leads kept track of the topics discussed in the meetings and any future actions arising out of the meetings through coaching records. In this way, both Research Leads and teachers were able to track any developments and improvements in the way teachers use the app and provide feedback to pupils. Coaching followed general guidelines, but the content and nature of the coaching was expected to be adapted depending on the needs of the teacher.

Evaluation objectives

The impact evaluation was designed to investigate the following hypotheses:

- Hypothesis 1: Digital feedback (including immediate and delayed) will have a positive effect on the outcomes (in mathematics) of pupils who are in randomly assigned intervention classes, compared to pupils in control classes.
- Hypothesis 2: Digital feedback (including immediate and delayed) will have a positive effect on the outcomes (in mathematics) for pupils eligible for FSM, compared to FSM pupils in control classes.
- Hypothesis 3: Digital feedback (including immediate and delayed) will have a positive effect on the engagement of pupils who are randomly assigned to intervention classes compared to pupils in control classes.

The evaluation protocol (Sim et al., 2018, Trial Protocol, Version 2, dated 11 September 2018) and Statistical Analysis Plan (SAP) (Sutherland and van Belle, 2018, SAP, Version 2, dated 10 October 2018) are available on the EEF website.[3]

Ethics and trial registration

The study was reviewed by the RAND Europe ethics advisory board and approved by the RAND Corporation Human Subjects Protection Committee (HSPC). This review included a requirement to obtain consent from teachers to

[3] https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/digital-feedback-in-primary-maths/

participate in the evaluation, and parental opt-out from the additional testing for the evaluation component of the project, with participants' parents or legal guardians being the decision-makers for individual pupils.

School recruitment was the responsibility of the KYRA Teaching School Alliance. Participating schools signed a Memorandum of Understanding (MoU) which outlined the roles and responsibilities of all stakeholders involved. The MoU made it clear that once schools agreed to participate, the expectation was that final outcome testing of pupils would be allowed, even if the school withdrew from the intervention. The MoU provided to schools is included in Appendix C.

The trial was registered under the following International Standard Randomised Controlled Trial Number (ISRCTN): ISRCTN50447665.[4]

Data protection

RAND Europe obtained personal data from pupils, teachers, and headteachers from the KYRA Teaching School Alliance as a data controller. Basic pupil information was obtained on the basis of legitimate interests from schools pursuant to data-sharing undertakings or agreements with the KYRA Teaching Alliance (see Appendix H for the data that was shared). The agreement outlines all obligations of partners, as well as a schedule outlining which data will be collected, how it will be shared, and how it will be kept secure. It clarifies that data collected may be held by RAND Europe until the end of this study.

RAND Europe obtained baseline data from the National Pupil Database and pupil outcome data from its subcontractor (Australian Council for Educational Research, ACER), which acted as a processor pursuant to appropriate data-sharing terms in its subcontract. Personal data obtained by ACER was on the basis of legitimate interests, and pupils and parents were provided with age-appropriate fair-processing privacy notices that explained the use, storage, and secure handling of the data. RAND Europe provided a privacy notice to schools (see Appendix D) and asked schools to share this with parents. This privacy notice explained the trial in short, and detailed what data was collected, how it was collected, why, what the legal basis for processing was, what the data was being used for, how it was shared between parties involved in the study, how it was being kept secure, and for how long it was going to be kept (end of trial; see Appendix D).

Data sharing between RAND Europe and ACER was governed by a contract between RAND Europe acting as Data Controller and ACER as Data Processor. A variation was made to this agreement in August 2018 to reflect and comply with changes in EU legislation (the introduction of the General Data Protection Regulation). This amendment clarifies the processing duties under GDPR, obliges both parties to confidentiality and to taking adequate security measures, and obliges partners and authorities to notify each other in case of data breaches. Further, it requires ACER to assist RAND with responding to requests made by participants of the study with regards to their data rights, and ensure that sufficient deletion procedures are put into place.

Furthermore, in order to obtain data from the selected case study schools for the impact process evaluation of the trial, RAND Europe set up data-sharing agreements with schools directly—with both RAND Europe and schools acting as Data Controllers (see Appendix G for a data schedule). This sets out the terms and conditions on which data will be securely shared by specific means between the partners and that it may be kept until this study concludes.

Project team

The intervention was developed and implemented by James Siddle for the KYRA Teaching School Alliance. As such, James Siddle and the KYRA Teaching School Alliance were responsible for the recruitment of schools and the training of teachers and Research Leads as well as implementation, including running the PLCs in association with staff from Bishop Grosseteste University. This collaboration provided the intervention team the capacity needed to implement the intervention. Those selected as PLCs coordinators worked as lecturers focusing on primary education at Bishop Grosseteste University, with some specifically focused on maths and/or information technology.

[4] https://doi.org/10.1186/ISRCTN50447665

The team in the KYRA Teaching School Alliance comprised James Siddle (project leader), Sarah Porter (administrator), and Lisa Yates (former administrator).

The evaluation was conducted by RAND Europe. RAND Europe was responsible for the overall design of the evaluation including outcome and process evaluations, trial design, analysis, reporting, and quality assurance of the evaluation. A testing subcontractor, ACER,[5] was responsible for recruitment, outcome testing, distribution of process evaluation surveys, and marking the maths outcome test.

The evaluation team in RAND Europe comprised Dr Alex Sutherland (project leader), Miriam Broeks (project manager), Eleftheria Iakovidou and Natalie Picken (field work and analysis), Dr Megan Sim, Dr Janna van Belle, Dr Julie Bélanger and Dr Joachim Krapels (formerly RAND Europe). The evaluation team also benefitted from advice and quality assurance from Dr Jon Schweig (RAND Corporation) and Dr Yulia Shenderovich (formerly RAND Europe).

[5] https://www.acer.org/gb/assessment

Methods

Trial design

Trial type and number of arms		Two-arm, cluster-randomised controlled trial
Unit of randomisation		Class
Stratification variable(s) (if applicable)		n/a
Primary outcome	variable	Mathematics attainment
	measure (instrument, scale)	ACERs Essential Learning Metric (ELM) for mathematics for Years 4 and 5
Secondary outcome(s)	variable(s)	Pupil disengagement
	measure(s) (instrument, scale)	Engagement versus Disaffection survey (Skinner et al., 2008)

The trial was a two-arm cluster-randomised controlled trial with classes in Year 4 and Year 5 as the unit of randomisation, that is, all Year 4 and Year 5 classes in each participating school were randomised into treatment or control condition. A school- or year-level randomisation would have been preferable to avoid risks of contamination across classes, but recruitment and delivery limitations meant this was not feasible. Specifically, the delivery team faced challenges recruiting enough schools to allow for school- or year-level randomisation (too few schools would have meant an under-powered trial). Further, there were also logistical limits on delivery that meant not as many schools could be included in the trial. In order for the trial to proceed, the compromise was to implement a within-school design, randomising classes. In schools where teachers teach a combined Year 4 and Year 5 class, this combined class was included as a single year-group in the randomisation and analysis.

For control classes, it was 'business as usual' during the school year 2017/2018. No incentives, wait-list or alternative treatment was offered as all schools in the trial received training in digital feedback. The aim of the efficacy evaluation was to assess whether Digital Feedback leads to an improvement in maths performance greater than that observed in ‘business-as-usual’ classrooms.

Participant selection

Schools were recruited by the KYRA Teaching School Alliance in Lincolnshire, North Lincolnshire, Nottinghamshire, and the Humber based on their capacity to recruit within these areas but not beyond. The following school eligibility criteria applied:

- only state schools were included;
- schools needed to have enough tablets for one year group/class (roughly one tablet for every four to five pupils); and
- schools should not have previously received training in the Digital Feedback intervention.

Only Year 4 and Year 5 teachers were included in the trial; all Year 4 and Year 5 pupils were eligible for inclusion. Parents were given the opportunity to withdraw their child’s data from the evaluation (see Data Protection section).

Outcome measures

Primary outcome

The primary outcome measure at the end of Year 4 and Year 5 was the ACER Essential Learning Metric (ELM) for Mathematics for Years 4 and 5. The tests were paper-based, took 45–50 minutes to complete, and were independently delivered, invigilated, and marked by ACER (please note that usual administration of ELM is online, and adaptive, but uncertainty about school IT systems meant that a paper-based differentiated test was created for the trial by ACER).

The ELM Mathematics test consists of two sets of measures: mathematical content (number, algebra, measurement, geometry, statistics, and probability) and mathematical processes (understanding, fluency, problem-solving, and reasoning). The test was reviewed by the KYRA Teaching School Alliance for compatibility with the mathematics curriculum taught in the test year (2017/2018). After administration of the test, data was supplied to RAND Europe for analysis by ACER in both its raw and scaled score format. As described in the analyses section below, the raw score was used for the analysis. Schools were not told in advance what the test was (that is, the name of the test) but were informed about the general aspects it would measure and that the test was aligned to the National Curriculum of England.[6] All test invigilators were blinded to treatment allocation.

Secondary outcome

For the secondary outcome, ACER also collected pupil disengagement survey data at the same time as administering the ELM Mathematics test. Increased pupil engagement was hypothesised to be one of the key mechanisms through which digital feedback would improve pupil outcomes (see logic model, Figure 1, page 7). Measuring pupil disengagement was therefore crucial to understanding the mechanisms underlying the intervention. The instrument used to measure pupil disengagement was the Engagement versus Disaffection survey developed by Skinner et al. (2009). The scale captures behavioural and emotional engagement and dissatisfaction in the classroom and has adequate score reliability (0.79), adequate cross-time correlation (0.63), and high factor loadings on the factors (ranging 0.55–0.84). Further details on the psychometric properties of the test are reported at length in Skinner et al. (2009).

Sample size

At protocol stage (see Sim et al., 2018, Trial Protocol, Version 2, dated 11 September 2018), the aim was for the trial to consist of 120 year groups over 60 schools with 60 year groups (30 in Year 4 and 30 in Year 5) in each treatment arm. Using these as starting points—and using data provided by KYRA and the EEF on the average number of pupils per school in the regions identified—we assumed that there are on average 25 pupils per class. We assumed one class per year group and two year groups (4 and 5) per school. With the assumed (average of) 25 pupils per class and 120 classes, the overall sample for the study was estimated to be approximately 3,000 pupils. We further assumed, based on EEF guidelines (EEF, 2018b) and requirements for power calculations, that the amount of variation explained by covariates is 0.53 for level 1 and 0.00 for level 2, and that the intra-cluster correlation coefficient (ICC) is 0.13 (13%). Lastly we also chose an alpha of 5% and an intended 80% power to detect effects. Power and minimum detectable effect size (MDES) calculations were performed using PowerUp! (Dong and Maynard, 2013).

At randomisation stage, the KYRA Teaching School Alliance had recruited 34 schools (or 67 year groups—one school had a split Year 4/5 class, which was treated as a Year 5 class in the analysis). Because the school and year group recruitment was below target, we decided (after discussion between RAND Europe, the KYRA Teaching School Alliance, and the EEF in May 2017) to change the randomisation plan from year-group randomisation to *class-level* randomisation. The presence of a few large schools in the sample meant that the relatively small number of schools

[6] The following information was provided to schools about the test: 'ELM's Mathematics is a mathematics assessment developed for schools in England. It measures what pupils know, understand, and can do in the learning area of mathematics. For this study, the tests are provided on paper. All questions are multiple-choice and children mark their answer by ticking the box next to their selected answer. While the content is aligned to the National Curriculum in England, there may be questions that require knowledge the pupils in your class have not yet been taught or questions relating to previous years' studies. This is intentional and designed to capture the breadth of skills and knowledge within each year group across the country'.

provided 108 classes for randomisation. With the assumed (average of) 25 pupils per class and 108 classes, the overall planned sample for the study was approximately 2,700 pupils.

At analysis stage the final sample of pupils was 2,564, but problems with linking allocation and teacher data to pupils (discussed below) meant the ITT analysis sample for the primary outcome was 2,133.

Randomisation

The unit of randomisation was Year 4 and Year 5 classes within schools. Schools were asked to provide teacher names for each class in the following academic year so that randomisation could take place.

Allocation to treatment or control was conducted on 19 June 2017 by Dr Sutherland and included all classes for which schools had signed the MoU. Recruitment pressures meant that not all schools had supplied baseline data (that is, linked pupil and teacher data) at the point of randomisation, but had indicated how many classes and the names of teachers who would be teaching a given class. We randomised based on the available information but the issue, discovered later, was that the information supplied did not match the later reality.

A stratified randomised design was planned, stratifying by year, with classes within Year 4 and Year 5 as the unit of randomisation. However, some recruited schools misunderstood what participation in the trial would entail in terms of random allocation; for example, that there was the possibility that a participating school would not be allocated a treatment class or that schools could not define which classes would receive the intervention. In the case of the latter, the delivery team communicated to schools that nominating a treatment class would not be possible. Nonetheless, given that some schools communicated the wish to have a treatment teacher in order to participate in the trial, this meant that there was a need to ensure that *each school* was allocated at least one treatment class. That in turn meant that randomisation stratified by year was not possible. To accommodate the request to ensure each school would have at least one treatment class, randomisation took place separately within each school. Due to the higher than anticipated number of small schools (schools with only one or two year groups, or single form entry), stratification within schools by year group was considered undesirable. Ultimately, classes were used as the unit of randomisation. In hindsight, we also believe that some schools may have wished to *only have* one treatment class and may have been less keen to participate when allocated more than one treatment class; however, whether this was in fact the case is not entirely clear to the evaluation team.

In practical terms, we assigned random numbers to the initial list of 108 classes that had been recruited by the delivery team. We then sorted the data by school and assigned the first (randomly ordered) class within each school to the treatment condition and the second to the control condition. Subsequent classes were then assigned to treatment and control in turn. These steps were undertaken using code run in Stata (StataCorp, 2017) so that although 'sequential', allocations were instantly completed. This approach assumed schools had equal numbers of Year 4 and Year 5 classes, to avoid unbalanced treatment allocation. In reality there were more schools with odd numbers of classes ('misfits'), which resulted in slightly more treatment classes than control groups. The evaluation team informed the KYRA Teaching School Alliance of allocations on the 19 of June 2017. Table 2 shows the allocations.

Table 2: Class-level allocations

Year	Control	Treatment	Total
Total	**52**	**56**	**108**

Issues with randomisation and baseline data created problems for analysis

As noted above, not all schools were able to supply pupil and/or teacher data at the point of randomisation. Those that did were informed of allocations straightaway by the delivery team, while others were not informed until later (when they had supplied pupil data). We know from the delivery team and through later contact with schools that the allocation of several classes to treatment created problems for schools, and some then withdrew from the intervention because of the perceived burden it would place on the school. For example, a three-form-entry school with two year groups would have three treatment classes, possibly split across two year groups. It seems that some schools were only expecting

one treatment class per school, and others simply reported they ‘didn't like the allocation’, which raises questions about how well those agreeing to participate actually understood what they were agreeing to.

The net result of the problems was that at the point of preparing for outcome testing in early 2018, it became clear there were some mismatches in terms of how much data we had been sent compared to how many classes/pupils were supposedly participating in the trial. As such, the evaluation and delivery teams, along with the testing company (ACER), coordinated to contact schools in order to request unique pupil number (UPN) data to link to National Pupil Database (NPD) data. This process resulted in the UPNs for 2,564 pupils being supplied and used for both the outcome testing and requesting baseline NPD data.

Defining the unit of randomisation (classes or teachers) and the need for additional data

The timing of the recruitment and class-level randomisation also presented issues for the evaluation. Recruitment took place between September 2016 and March 2017, many months before schools typically have determined exactly which pupils will be in which class, and which teacher will be teaching them. This uncertainty was one reason that schools found it difficult to supply pupil data (or be able to say which teacher would be teaching a particular class). It also raised a question about what, exactly, was being randomised at 'class level'. With hindsight, the messaging about this should have been that it was the *teacher* being randomised as this was a CPD-driven intervention focused on teachers meaning that those pupils that intervention teachers then went on to teach in the coming academic year would become intervention participants. In practice, schools sent us data for pupils that 'might' be taught by a specific teacher—this was particularly the case for large schools—and this had knock-on effects later on in the project (described in section ‘Participant flow including losses and exclusions’ on page 26).

Problems with data that affect the trial results

There were several challenges with the analysis for this trial that relate to randomisation, attrition, and the final analysis. We set these out in bullet points below, and then discuss them in more detail.

- Schools did not provide pupil data at baseline in time for randomisation. We did not reveal allocation data until we had received pupil data in these cases.
- Some schools provided pupil data after randomisation for more (or fewer) pupils than the number of classes randomised would suggest. For example, data for 90 pupils was eventually supplied, with only two classes randomised in a given school. In this situation we only used data for pupils who were originally included in the trial and deleted the additional data.
- Schools could not provide class lists with prospective teacher names associated. That is, current teacher names were often given (for the academic year when recruitment was happening), but it was often impossible for schools to accurately forecast which teacher would teach a given class the following year.
- When teacher names were given, these did not always correspond to class-identifying information (for example, a pupil could be in a specific class but the supplied teacher details were not linked to a class).
- The largest schools in the study were more likely to not supply teacher names at all. The three largest schools, contributing more than 400 pupils, did not supply teacher names that could be linked to pupils/allocations meaning that all of their data was excluded from the main analyses (n = 431 pupils). As this meant that both treatment and control classes were excluded in these schools, this lessens the threat to internal validity.

Taken together, we had pupils/classes missing teacher names at recruitment, uncertainty about which teachers taught which class in the study year, and in some cases were not sure about which years/classes pupils were in, and thus whether they were treatment or control conditions (or if treatment teachers actually did the intervention). As such, linking pupils, teachers, and classes with treatment-control assignment and baseline/outcome data was a significant challenge for the evaluation. To overcome this, through October 2018 to February 2019 we:

- asked schools for which we were not sure about allocation to provide UPNs for all pupils taught by specific teachers;
- reviewed all intervention logs collected by PLCs to record if/which teachers from which schools attended PLC meetings (encompassing five spreadsheets each logging attendance for the eight scheduled PLC meetings for

each teacher/school; it also included year group taught by teachers in all but one PLC record); when attendance at PLC meetings matched presumed allocation, we did not investigate further;

- where it was not possible to match teacher attendance at PLC meetings to teachers' allocations, we contacted these schools individually to ask which teachers were actually 'treatment' or 'control'—that is, whether they followed allocation—and to verify pupil class membership by checking small samples of pupil data against those in a given teacher's class; any changes of teacher for a given class were also recorded;
- re-checked treatment-control assignment using both the original data supplied by schools for pupil-teacher combinations and PLC records, which allowed us to confirm that in some instances all Year 4 pupils were treatment, and Year 5 were control (or vice-versa); and
- finally we rebuilt the entire pupil-level database and manually checked all pupil, year, class, teacher, and treatment group combinations against multiple sources (original data supplied by schools, PLC records, records from school contacts, and NPD data).

It is important to note that datasets were cleaned and recoded prior to linkage to allocations, so cleaning/recoding decisions were not biased by knowing allocations. Given the various steps above, we, of course, recognise that there are limitations with such exercises, least of all the various decisions that have to be taken. We logged these decisions and those informed our approach to sensitivity analysis. Without this effort we were only confident of allocations for 1,300 pupils after the third step, but after the fourth step, around 2,000 pupils could be more confidently linked to treatment/control conditions that linked back to the original allocations at class level. That translated into final samples of 2,133 pupils (n = 112 'classes') for the ITT analysis for the maths outcome and 2,063 (n = 110 'classes') for the FSM subgroup analysis, with the largest losses coming from the three largest schools.

Statistical analysis

Primary intention-to-treat (ITT) analysis

The analysis of primary outcomes (as detailed above) for both Year 4 and Year 5 pupils was on an ITT basis.[7] This method compares outcomes means for the treatment and comparison groups, and subjects are analysed according to their randomised group allocation regardless of whether or not they adhered to this group. The ITT approach is inherently conservative as it captures the averaged effect of *offering* the intervention, regardless of whether or not the participants comply with the assignment. Problems of drop-out/non-attendance may be an issue for this trial depending on how motivated teachers are (see Attrition section).

The unit of analysis was the pupil, for those in classes randomised to use Digital Feedback (or not). Following our SAP (Sutherland and van Belle, 2018), we accounted for clustering of pupils by class via multilevel models—but also used robust standard errors because we have classes clustered in schools as well as pupils clustered in classes. Multilevel models assume that higher level units in the study are a random sample of all units—which is sometimes a source of contention—but one benefit of this approach is being able to explicitly partition variance and more flexibly handle complex variation (Snijders and Bosker, 2012).

The general equation for the multilevel model is given below:

$$y_{ij} = \propto + X_{ij}\beta + Z_j b_j + \delta DF_j + u_{ij} + u_j \qquad i = 1..N, j = 1. \qquad (1)$$

Where y_{ij} denotes the pupil level outcome; i and j denote pupil and class indexes respectively; X_{ij} is the baseline attainment data. DF_j is a dummy variable denoting treatment/control group at class level; Z_j indicates class characteristics, in this case the year-group variable.[8] Finally, u_{ij} is the pupil-level error term, and u_j is the class-level error term. In equation (1) α denotes the overall model intercept. u_j are referred to as class-level residuals ($u_j \sim i.i.d\ N(0, \sigma_u^2)$) and e_{ij} are individual-level residuals ($e_{ij} \sim i.i.d\ N(0, \sigma_e^2)$). The coefficient δ associated with the DF_j

[7] https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Research_Report/2015_Analysis_for_EEF_evaluations.pdf

[8] In the SAP we described stratifying on the year group variable but that was not done, which led to an imbalance in year groups allocated to treatment and control. As such, year group was included in all analysis models.

dummy constitutes the main result of the trial. ‘All pupils’ represent the main analysis, so to obtain the main result of the trial, Equation (1) above was estimated for all pupils.

Pre-intervention mathematical ability was captured using KS1 maths data from the National Pupil Database. The baseline pupil measure included in our analysis was a continuous variable for attainment, which is operationalised as Key Stage 1 prior attainment results in maths (the variable in the NPD is KS1_MATPOINTS). These were collected either one or two years before randomisation (for Year 4 and Year 5 pupils respectively). This pupil data was obtained from two sources. Initially, and prior to randomisation, from schools: lists of pupils in the target year groups were provided by all in-trial schools before schools were made aware of how their classes were allocated to control or treatment conditions. All analyses were performed in Stata, version 15 (StataCorp, 2017).

The difficulty faced regarding the baseline data was that the measure of KS1 maths changed between the two year groups (ignoring the maxim, 'if you want to measure change, don't change the measure' Hout, no date). This meant we were faced with either treating all KS1 data as categorical, that is, including n-1 dummies for all levels / categories, or find an alternative (such as KS1 data from English to include some measure of pupil ability). In the end, the evaluation team opted to create a dummy variable that recorded whether the pupil was at/above the expected level for their respective cohort/testing regime (0 for ‘no’, 1 for ‘yes’). Whilst arbitrary, this seemed less problematic than other solutions. The effect of this was that the majority of pupils in each year group were considered to be 'at or above' the required level.

Subgroup analyses

Acknowledging that the trial was not designed/powered for an FSM subgroup analysis, the evaluation team also conducted a subgroup analysis for ‘Ever FSM’ pupils, following the EEF’s approach for FSM subgroup analysis (EEF, 2018b). The impetus for this is the previous evaluation findings that suggested stronger results for FSM pupils (Siddle, 2015). While we believe that this option has the potential to reveal additional, indicative evidence, we acknowledge that this particular study is not sufficiently powered to result in strong conclusions about the efficacy for FSM-eligible pupils.

Interim analyses

No interim analyses were planned or conducted.

Imbalance at baseline for analysed groups

A well-conducted randomisation will, in expectation, yield groups that are equivalent at baseline (Glennerster and Takavarasha, 2014). Because classes within schools here were randomly allocated to the control and intervention conditions, any imbalance at baseline would have occurred by chance. To assess imbalance at baseline, we compared groups at class and pupil level. In addition, given that class-level randomisation was used, and in accordance with the SAP, the evaluation team felt it was necessary to check some teacher-level variables as an additional balance-check procedure. Therefore, the following variables were examined by means of cross-tabulations and histograms that assess the distribution of each characteristic within the control and treatment groups:

- teacher experience (years of experience);
- teacher’s age; and
- teacher’s gender.

These variables were chosen given the influence that teacher characteristics can have on their teaching quality and ultimately on pupil outcomes. There is broad consensus that ‘teacher quality’ is one of the most important school variables influencing student achievement (OECD, 2005).

At the pupil level, baseline balance was assessed for the following characteristics:

- age (year group);
- eligibility for FSM;
- gender; and
- KS1 attainment.

Statistical significance tests were not carried out to assess the balance as their premise does not hold in randomised controlled trials.[9] Instead, tables of the class-means (and standard deviation, where appropriate) for each characteristic are presented, and the magnitude of any differences explored.[10] For skewed variables we used quartile-based measures.

Secondary outcome analyses

Pupil disengagement was measured using the Engagement versus Disaffection survey (EvDwLq) (Skinner et al., 2008). We compared treatment and control classes using the same model specification as above, substituting the secondary outcome, and without baseline attainment in the model as baseline scores are not available. The EvDwLq has positively- and negatively-worded questions, but the summed scale measures the overall level of disaffection with school—as such, higher scores mean a greater level of disaffection.

Effect size calculation

With the multilevel models, we used the effect sizes (ES) for cluster-randomised trials given in the EEF evaluator guidance. An example, adapted from Hedges (2007), is given in equation (2):

$$ES = \frac{(\bar{Y}_T - \bar{Y}_C)_{adjusted}}{\sigma^2} \qquad (2)$$

Where $(\overline{\boldsymbol{Y}}_{\boldsymbol{T}} - \overline{\boldsymbol{Y}}_{\boldsymbol{C}})_{\boldsymbol{adjusted}}$ is the mean difference between intervention groups adjusted for baseline characteristics and σ^2 is the population standard deviation (variance), which we requested from the Department for Education. The ES therefore represents the proportion of the population standard deviation attributable to the intervention (Hutchison and Styles, 2010). A 95% confidence interval for the ES, that took into account the clustering of pupils in classes, is also reported and was calculated by dividing the upper and lower confidence interval bounds by the population standard deviation.

As the population variance was not available we substituted the sample-based, pooled unconditional variance (s*) in the denominator, which is an estimate of the population variance. In multi-level models this variance would have been the weighted average of the variance of treatment and control groups, as per EEF guidance (EEF, 2018b: 4 footnote 11). The formula for the denominator is given in equation (3):

$$s^* = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \qquad (3)$$

Missing data

Missing data can arise from item non-response or attrition of participants, teachers, or schools. We first determined the proportion of missing data in the trial. Our use of administrative data for pupil baseline data reduced missingness arising from both item non-response and attrition.

Our analysis would always begin with an ITT approach as set out above, meaning that we would look to analyse all classes and pupils as allocated at randomisation, regardless of compliance, and then move from there depending on the extent of missingness and where it occurs in the data.

We intended to explore attrition across trial arms as a basic step to assess bias (Higgins et al., 2011). Below we provide cross-tabulations of the proportions of missing values on all baseline characteristics (as detailed in the previous section, at both pupil and school level), as well as on the primary outcome measures.

Our plan for assessing missingness also included modelling missingness at follow-up as a function of baseline covariates, including treatment. However, the problems with linking allocation data meant this was not sensible,

[9] http://www.consort-statement.org/checklists/view/32-consort/510-baseline-data

[10] There is a convention in some disciplines that a 10pp (or larger) difference in treatment and control means at baseline constitutes 'imbalance' is thus justification for including those measures in sensitivity analyses.

furthermore, there was no loss to follow-up on the outcome. We discuss how we dealt with missingness on allocation data in the 'Attrition' section (page 26), and report sensitivity analyses where we assigned all missing allocation data to treatment and control. The inherent issue is that because it was classes allocated rather than schools, it means that in some cases we are merely mis-assigning half of all classes in the school each time. In other words, if there were three control and three treatment classes in a school missing allocation data, assigning all classes to treatment/control would mean in each analysis half were wrong each time (we considered also randomly assigning pupils but unless this was done iteratively we cannot see how this would lead to any further useful information). Given that both treatment and control classes were missing from schools, overall missingness was similar in intervention arms, and missing data on the allocation data would be unrelated to final test scores, risk of bias should be minimised.

Non-compliance with intervention

According to the Digital Feedback development team, successful implementation of the intervention depends on several factors:

- Fidelity and dosage: Are teachers implementing the intervention as intended according to the Research Lead?
- Quality: Are teachers aware of, and providing, feedback that adheres to the effectiveness quality according to the Research Lead?
- Targeting: Are teachers using the diagnostic tools to target delayed feedback at pupils with identified needs according to the Research Lead?

In the end, we agreed that the compliance metric for analysis would focus on fidelity and dosage measures to operationalise compliance with the intervention (see below), whereas factors 2 (quality) and 3 (targeting) would be analysed as part of the process evaluation (Sutherland and van Belle, 2018).

We had planned to collect details of compliance via three different methods in order to triangulate the data and create a compliance measure. However, there were issues with collecting some of this data that we describe briefly next.

Data from the digital feedback app

The intention was to use metadata from the app, specifically data that showed how often the app had been accessed by the teacher (a date-timestamp and a record of file size). For each class, we had planned to use the average number of times the app had been accessed per day (counting term-dates only) as a rate per 25 pupils (so that schools with smaller or larger classes are standardised). However, in practice it was not possible to collect metadata from the app as not all teachers used the app (see further discussion in Implementation and Process Evaluation).

The plan was to derive a Complier Average Causal Effect (CACE) measure by using app data. If the app data was available, intervention classes would be counted as compliers if they had accessed the app at least once a week for at least 30 out of 39 weeks in the school year. However, we had specified in the SAP (Sutherland and van Belle, 2018) that if we were unable to get access to app-data, we would use the two alternative measures discussed below.

Data from weekly coaching logs/records

Coaching records on the use of Digital Feedback in the classroom (self-reported utilisation) were going to be provided by teachers to their PLC coordinators who would pass this information to the intervention team in their Red/Amber/Green (RAG) ratings. 'Compliance' was going to be defined as the use of diagnostic assessments *at least once a week* for at least 30 out of 39 weeks in the school year. However, after checking with the implementation team, it was confirmed that PLC RAG ratings did not systematically collect information about coaching records (that is, on the use of Digital Feedback in the classroom or on how often coaching sessions took place). Furthermore, PLC RAG ratings we received did not contain systematic information on the frequency diagnostic assessments were used. Despite obtaining a sample of coaching records from case study schools, no robust compliance analysis on the use of diagnostic assessments could be conducted based on these.

Self-reported usage of the app by teachers.

One of the questions in the teacher survey was, '*How often do you use the following two types of digital feedback? Once a day or more than once a day / several times per week / once a week / once every two or three weeks / once a month?*'—for both immediate and delayed feedback. The plan was to count teachers as compliers if teachers indicated that they had used digital feedback at least once a week both in the midline and the endline survey. However, despite the survey being sent to all intervention teachers, response rates were low and declined between surveys, making this measure inaccurate to calculate compliance.

The most complete data available that could have been used to analyse compliance were PLC records (attendance of teachers to PLC meetings). While these records revealed that some schools did not attend any of the PLC sessions, they were in some cases incomplete (in two cases only information up to PLC session 3 and 4 were included; in one case only school names rather than teacher names were logged). Given these issues, PLC records could not be used as a strong enough data source to conduct compliance analysis. Furthermore, we were unable to triangulate survey with PLC data due to surveys being anonymous and not representative of all participants. Had data been more complete, triangulating these two data sources would have served as a good approach to provide a more complete picture of compliance. Nonetheless, data presented in the 'Implementation and process evaluation results' section suggest that compliance with the intervention requirements was low.

Implementation and process evaluation (IPE) tasks

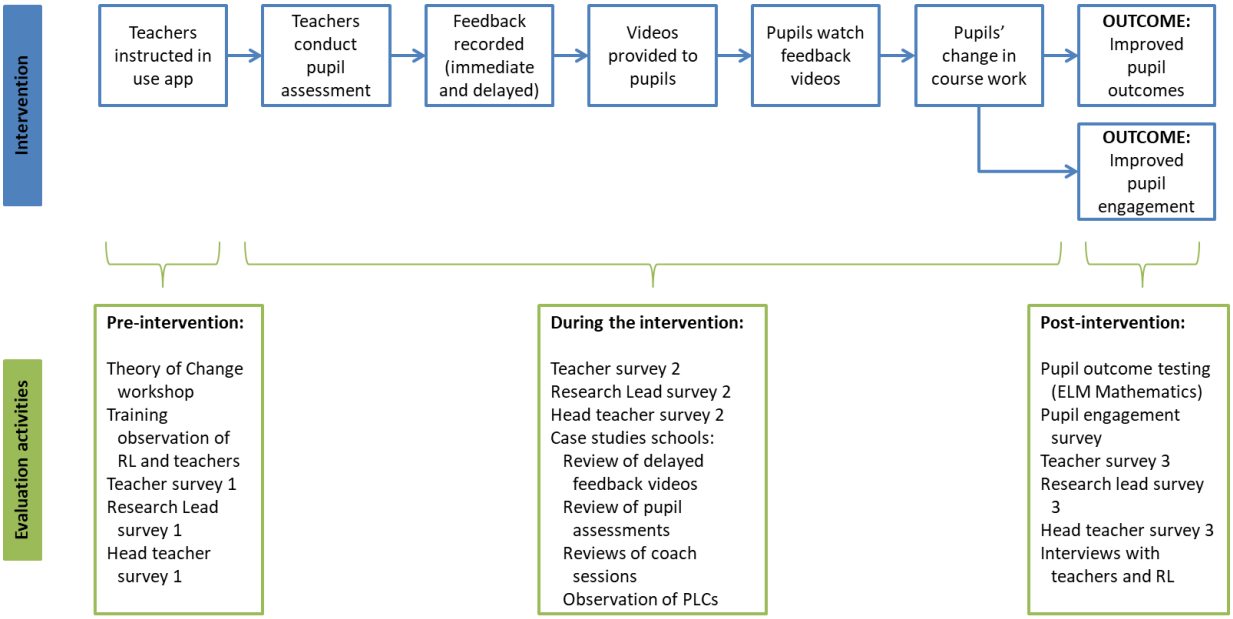
Following Oakley et al. (2006) and the EEF guidelines to process evaluation (Humphrey et al., 2015)—and informed by implementation science and realist approaches to evaluation—we conducted an implementation and process evaluation to better understand the complexities of the Digital Feedback trial. Implementation science emphasises thinking about 'dosage' or 'exposure', fidelity/adherence, and implementation quality. The realist approach emphasises thinking about what works, for whom, and under what conditions. The purpose of the process evaluation was to address the following questions:

- Was the intervention implemented with fidelity in the intervention schools?
- What factors and initial conditions appear to explain variation in fidelity of implementation?
- What appear to be the necessary conditions for success of the intervention?
- What were the barriers to delivery?

The process evaluation aims to examine the mechanisms of the intervention and informs the interpretation of findings from the quantitative analysis. Figure 2 below illustrates the methods we used for the implementation and process evaluation. For the process evaluation, we collected data at three points: pre-intervention, during the intervention, and post-intervention. In this section we describe the various tasks planned and undertaken for the process evaluation. The results from these activities are reported in the section '

Implementation and process evaluation'.

Figure 2: Three phases of implementation and process evaluation data collection



Pre-intervention IPE tasks

Theory of change workshop

The evaluation started with a theory of change (TOC) workshop in which representatives of the EEF, Digital Feedback, and RAND Europe developed the intervention logic, including specifying in detail the mechanisms by which Digital Feedback was expected to affect pupil outcomes. The development of the intervention logic drew on the expertise of the intervention team and on a review of the existing literature by the evaluation team. The TOC describes how, and under what circumstances, Digital Feedback was expected to yield improved pupil outcomes. This is presented in **Error! Reference source not found.Error! Reference source not found.** above.

Observation of teacher training in digital feedback

The training of teachers and Research Leads in the use of the app to provide Digital Feedback, and conversely teachers and Research Leads' understanding of the app, is crucial for the success of the trial. Through attendance at, and observations of, five randomly selected training sessions, potentially important differences in the instruction in the use of the app and the delivery of Digital Feedback could be noted. A semi-structured protocol was used to capture whether (a) the training covered all key elements of Digital Feedback and (b) provided information on implementation requirements for the trial; in addition it sought to (c) qualitatively assess the levels of disengagement of participants and (d) the level of preparation of trainers.

First teacher online survey

At the beginning of the school year, we asked all teachers to complete an online survey on the following topics:

- basic teacher-level demographic information (if not available by other means), including gender and years of experience—in the school and in total;
- reported experience using technology in their teaching, including for providing pupil feedback;
- views on written versus digital feedback; and
- perceptions of the pedagogical value of written and digital feedback, whether there are differences in appropriateness of digital feedback depending on subject matter (maths or English), and the reaction and receptivity among pupils to written and video feedback.

Please note that questions about teachers' experience of the training were asked in the midline survey since the baseline survey sought to serve as a measure (at baseline) of differences or similarities between intervention and control teachers' views of the elements described above.

First Research Lead online survey

Several Research Lead online surveys were conducted during the school year to capture the views of Research Leads on the provision of Digital Feedback by teachers to pupils. In the first survey, Research Leads were asked about their role within their school and their views on effective feedback.

First headteacher online survey

Headteachers at intervention schools were also asked to complete an online survey to capture their perceptions of:

- digital technology in the classroom;
- the use of Digital Feedback by teachers; and
- the selection of the research lead in their school.

Intervention phase IPE tasks

Second teacher online survey

The second teacher online survey was conducted around the mid-point of the trial in February 2018. It focused on teachers' ongoing experience with the provision of Digital Feedback and the use of the resources to assess pupils' understanding of a topic. In addition, it focused on the experience of the teacher with the training they received, their perception of coaching, and their perception of the PLCs.

Second Research Lead online survey

Research Leads were expected to frequently meet with, and potentially observe, teachers using the app to provide Digital Feedback. As such, Research Leads were asked to give their view on how well and how often Digital Feedback was provided by teachers.

Second headteacher online survey

Headteachers at intervention schools were asked about their perceptions of:

- the process by which Research Leads were selected;
- the cost of the trial in the start-up phase;
- the use of Digital Feedback by teachers; and
- the awareness among teachers of the characteristics of effective feedback.

Observation of PLCs

Members of the evaluation team attended and observed the workings of three randomly-selected PLC sessions. The observations sought to gain insights into the quality of PLC meetings (alignment with Digital Feedback), as well as insights into the day-to-day running of the intervention and any questions or issues participating teachers might bring forward. A semi-structured observation protocol was used to assess the meetings' consistency with Digital Feedback guidelines and fidelity to implementation and evaluation requirements (for example, to capture comments around minimising contamination). The protocol also captured information on training quality (that is, the preparedness and level of knowledge of the PLC coordinator) and overall levels of engagement.

Focus group

A fourth PLC meeting was intended to be observed. However, the PLC coordinator was unable to attend and the session was cancelled soon before the start of the session. As a result the researcher present took the opportunity to collect qualitative data from teachers via an impromptu focus group (please note this was not specified in the trial protocol).

Six teachers were present in this session. Questions around their experience using Digital Feedback (enablers and barriers), the perceived usefulness of PLC meetings, coaching sessions with Research Leads, and awareness of evaluation requirements were discussed.

Reviews for six case study schools

For a sample of six schools we conducted a number of further in-depth reviews of Digital Feedback. The selection of the schools took into account a range of factors including size of school (small versus larger) and location (to ensure geographical spread) as well as school willingness and ability to participate as a case study school. The review for each school focused on the following three data sources: delayed digital feedback videos, diagnostic assessments, and coaching records.

Characteristics of schools selected for case study review

Table 3 presents the characteristics of the six schools selected for case study review as well as the data collected from these schools. Despite the headteacher of school C agreeing to take part in the case study and to return a data sharing agreement the school did not upload materials despite multiple follow-ups by the evaluation team. In the case of school E, following initial communications with the Research Lead we were unable to obtain a response for the school to set up a data sharing agreement. In the case of school F, the Research Lead was unable to upload the digital feedback videos due to IT issues at the school. However, this Research Lead provided written feedback explaining difficulties experienced by their school in implementing the intervention.

Table 3: Characteristics of schools selected for case study review

School	Location	Size	Case study data collected
A	Lincolnshire	Single form entry	24 feedback videos; 13 coaching records
B	Lincolnshire	Multi-form entry	15 feedback videos; 14 coaching records; 58 diagnostic assessments
C	Derby	Multi-form entry	-
D	North Lincolnshire	Single form entry	7 coaching records; 10 diagnostic assessments
E	Rotherham	Single form entry	-
F	Doncaster	Multi-form entry	-

Review of delayed digital feedback videos

Digital Feedback aims to provide not only more timely feedback, but also more relevant feedback for pupils. We developed a coding protocol to assess the quality of delayed feedback videos. This protocol was based on the guidance to effective feedback provided in the training sessions for teachers and was complemented with existing evidence on elements that lead to effective feedback (Australian Institute for Teaching and School Leadership, 2017). Reviewers assessed whether the following elements were present in the digital feedback videos:

- positive elements;
- constructive advice;
- changes in performance;
- self-assessment;
- task-specific;
- process-specific; and
- metacognitive feedback.

Detailed descriptions of the above areas are provided in Appendix E. We requested digital feedback videos from the six case study schools; however, only two schools provided videos. The video coding protocol was applied systematically to a random sample of ten delayed feedback videos received from these two schools. Reviewers coded each element of the ten videos as ‘1’ if the element was observed, ‘0’ if not, or ‘N/A’ if not applicable.

Review of the diagnostic assessments

Although the evaluation team sought to analyse a selection of the diagnostic assessments used in the classroom, we only received a small number of assessments from two case study schools (see Table 3). Staff in three case study schools reported that it would be cumbersome to scan all of the diagnostic assessments, some of which were in pupils' textbooks. To assess the function the diagnostic tests perform in tailoring/targeting the digital feedback, the evaluation team would have needed to pair diagnostic assessments with the associated feedback video. However, this was not possible since schools were unable to provide diagnostics tests alongside their associated video feedback, in some cases due to videos being wiped from iPads on a regular basis. Therefore, the diagnostic assessments were not included in the analysis for the implementation and process evaluation.

Coaching records

After each weekly coaching session Research Leads completed a short summary form ('log'). For the six selected schools, we intended to review a selection of the forms to explore:

- Fidelity and dosage: Are teachers implementing the intervention as intended according to the Research Lead?
- Quality: Are teachers aware of, and providing, feedback that adheres to the effectiveness quality according to the Research Lead?
- Targeting: Are teachers using the diagnostic tools to target delayed feedback at pupils with identified needs according to the Research Lead?

With these questions in mind—and also with reference to the training teachers and Research Leads had received at the beginning of the intervention on effective feedback—we prepared a structured coding protocol to review the coaching records. This protocol was developed based on the guidance in the Digital feedback training materials. Three case study schools provided their coaching records (as presented in Table 3 above). The coaching record coding protocol was used for all the coaching records that were received from case study schools (protocol used is included Appendix E).

Post-intervention IPE tasks

Third teacher online survey

After the intervention, all teachers received a final online survey to capture their experience over the previous school year. In addition, teachers in intervention schools were asked about their views about using the app. The topics of the survey covered the same content as the first teacher survey:

- reported experience using technology in their teaching, including for providing pupil feedback;
- views on written versus digital feedback; and
- perceptions of the pedagogical value of written and digital feedback, the differences between subjects for which video feedback would be useful, and the reaction and receptivity among pupils to written and video feedback.

Third Research Lead online survey

The third Research Lead survey was also conducted after the intervention has been completed. Similar to the first Research Lead survey, the third survey focused on the perceptions of the Research Lead on the use and implementation of Digital Feedback among the teachers they coach.

Third headteacher online survey

In the third headteacher survey, all headteachers were asked for their perceptions on the implementation of the intervention. In addition, headteachers were asked to estimate the costs of running the intervention in their school.

Interviews

A selection of teachers involved in delivering Digital Feedback at the six case study schools identified above were also invited for semi-structured in-depth telephone interviews to get a detailed understanding of their experiences in using the app (enablers and barriers) and on their overall views on the intervention's different components. From the three teachers that were approached to participate in an interview, only one agreed to participate (the target was to interview six teachers, however, we were unable contact three teachers since their Research Leads either declined participating in the case study or were unresponsive to our communications). Furthermore, it had been originally planned to also interview Research Leads. However, given the additional requests made to them—to provide materials for the case studies (upload feedback videos, diagnostic assessments and coaching records)—it was decided not to approach Research Leads to reduce the burden of evaluation tasks on them.

IPE analysis approach

As part of the analysis of IPE data, we initially analysed each data source separately (surveys, training and PLC meetings observations, case studies, coaching records, digital feedback videos, and interviews) and then compared them to draw out emerging themes. Analyses of survey data focused on looking for any emerging patterns in responses, while bearing in mind methodological limitations (such as the inability to track an individual's responses across surveys). Surveys provided the most comprehensive data source as they covered the largest number of views across different participant groups (teachers, research leads, and headteachers). Surveys also covered questions on the training, early set-up, and implementation of the intervention as well as questions around daily use (enablers and barriers). Analyses of this data were complemented by information obtained through training and PLC observations, and case study data. By triangulating different data sources the evaluation team was able to obtain a more comprehensive picture of the implementation and use of the intervention in practice.

Costs

We estimated average marginal costs per pupil per year for schools for the intervention in two stages. First, following an initial discussion with the intervention developer, we asked headteachers to report start-up and running costs using a template outlining key cost categories in both areas. Average marginal costs per pupil were estimated in line with the convention followed by other Education Endowment Foundation projects (EEF, 2016). The EEF cost rating is included in Appendix A. Second, separate estimates of prerequisite costs and of additional staff time were also requested from school staff. These estimates together provide a general indication of the costs involved with implementing the Digital Feedback research, though we recognise that costs may differ between schools; in particular prerequisite costs may vary depending on existing resources available.

We collected data on the following specific cost categories:

- direct financial costs, including direct costs paid for tablets and salary costs for additional staff required to conduct training and coaching;
- direct time costs, including staff time and volunteer time; this will be reported separately—we will also report average supply cover time (rather than cost) required for the intervention; and
- prerequisite costs (for example, tablets).

We collected additional data from the intervention team to validate the above estimates and to ascertain whether any changes to costs are anticipated in the future if the intervention were to be made more widely available.

Timeline

Table 4 below presents a timeline of the main activities related to the evaluation and intervention delivery.

Table 4: Timeline

Date	Activity
Sept–October 2016	Evaluation kick-off meetings
Sept 2016–March 2017	Recruiting schools and teachers
May–June 2017	Randomisation
May–July 2017	Training provided for teachers and research leads; observations of training
September 2017	Intervention begins; teachers to start using Digital Feedback; baseline teacher (intervention and control) / Research Lead / headteacher surveys
February 2018	Mid-intervention teacher/ Research Lead / headteacher surveys in intervention schools
June 2018	Outcome tests for all pupils
June–July 2018	Follow-up survey with all teachers /Research Leads / headteachers
July–November 2018	(Phone) interviews with teachers and Research Leads in intervention schools
Spring 2019	Publication of the EEF report

Impact evaluation

Box 1: Impact evaluation—key findings

- The over-arching finding from the impact evaluation was that, as implemented in this trial and compared to 'business as usual', Digital Feedback did not have any discernible impact on pupils' maths ability. This result was robust to different specifications and assumptions. The same result also applied for pupils eligible for FSM.
- We also assessed the impact on pupil disengagement, which was proposed as a putative mechanism through which attainment might increase. We found consistent, but non-significant, results showing that treatment pupils reported being more engaged that control pupils. This results was also robust to specification changes and different assumptions, but was not substantially different for FSM and non-FSM pupils.

Participant flow including losses and exclusions

Figure 3 sets out the participant flow through the project. The reader will note that in places the number of pupils in treatment and control groups is not stated. As we set out below, there were a number of problems with data in the project that meant we were unable to reliably link pupil data with allocations. This is in spite of the post-test data collection being completed for all pupils in the trial. In our analyses we focus on those pupils we have been able to link allocation data for, but incorporate those missing this information into sensitivity analysis.

The number of eventual 'classes' in the analysis was also higher than randomised, but that was a function of assigning a 'missing' indicator for ‘teacher’ to pupils in a school where the treatment/control allocation was known. In a simple example, in a school with four classes but teacher names for only three classes and allocation data known for all four classes, there would be a 'missing' teacher; those pupils without a teacher would be grouped together for analysis. In total, there were 27 groups of pupils without teacher names but school and allocation data were available. As such, we included the allocation, school name, and 'teacher ID' to group pupils, meaning that only pupils with treatment, school ID, and teacher ID (even if teacher ID was 'missing') were coded for inclusion in the final analysis. Including treatment allocation meant we ensured separation between treatment and control classes, even where teacher names were missing. Note that the number of classrooms in treatment/control reported below differs from that in the SAP. Having checked the original allocation file (an Excel file locked for editing), the SAP contained a typographical error. The original treatment/control split is as reported below (56 treatment: 52 control).

At the point of analysis, the problems with teacher ID meant that the final analysis was conducted on 112 'groups' of pupils, the majority of whom were in *known* classes and/or year groups based on teacher allocation following class-level randomisation. However, for some pupils we categorised them based on other available data that enabled inferring their allocation. For example, if there were two classes in a school and the allocation was known (that is, we know from other data that one class is treatment and the other control) but no teacher names were known, these two classes were categorised as two different groups despite not having teacher names attached to them. To ignore this would have meant that all pupils from that school would have been considered part of the same class, which would lead to mis-estimated standard errors because the extent of clustering would be incorrect.

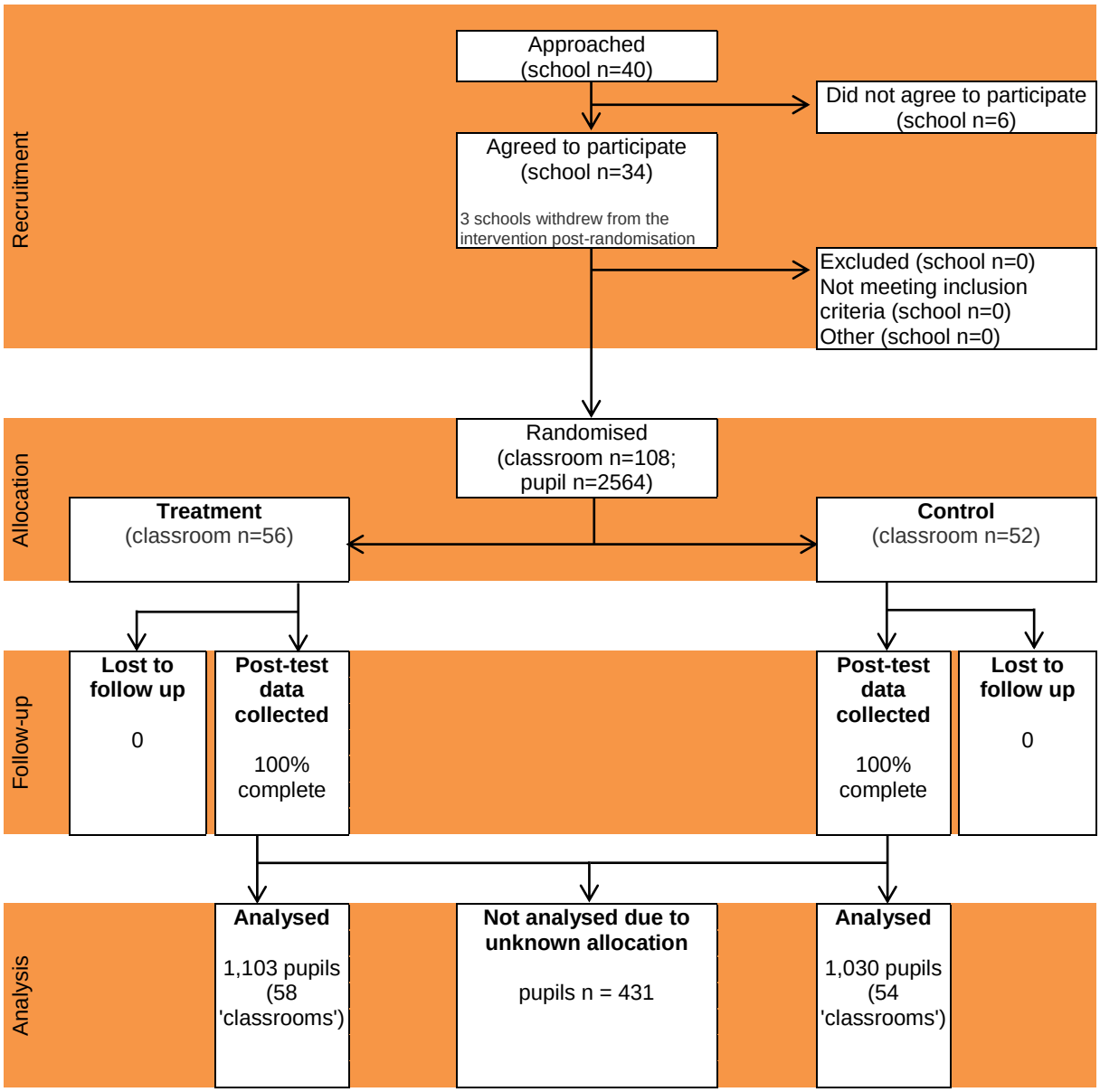
Table 5 presents the minimum detectable effect size at different stages of the trial.

Table 5: Minimum detectable effect size at different stages (ELM Mathematics)

		Protocol		Randomisation		Analysis	
		Overall	FSM	Overall	FSM	Overall	FSM
MDES		0.186	n/a	0.187	n/a	0.183	n/a
Pre-test / post-test correlations	level 1 (pupil)	0.53	n/a	0.53	n/a	0.53*	n/a
	level 2 (class)	0	n/a	0	n/a	0	n/a
	level 3 (school)	n/a	n/a	n/a	n/a	n/a	n/a
Intracluster correlations (ICCs)	level 2 (class)	10%	n/a	10%	n/a	9%	n/a
	level 3 (school)	n/a	n/a	n/a	n/a	n/a	n/a
Alpha		0.05	n/a	0.05	n/a	0.05	n/a
Power		0.8	n/a	0.8	n/a	0.8	n/a
One-sided or two-sided?		Two	n/a	Two	n/a	Two	n/a
Average cluster size		25	n/a	24	n/a	19.0	n/a
Number of classes	intervention	54	n/a	56	n/a	58	n/a
	control	54	n/a	52	n/a	54	n/a
	total	108	n/a	108	n/a	112	n/a
Number of pupils	intervention	1,350	n/a	1,282	n/a	1,030	n/a
	control	1,350	n/a	1,282	n/a	1,103	n/a
	total	2,700	n/a	2,564	n/a	2,133	n/a

* Assumed as the baseline variable was a three-level categorical variable.

Figure 3: Participant flow diagram



Attrition

We had collected outcome test data for 2,564 pupils in 34 schools (100% of those randomised). As set out in the preceding section, the study suffered several problems in terms of data linkage so as we were not confident of treatment-control group membership for 431 pupils, we excluded them in our final analyses (meaning 83% of pupils were retained for the primary outcome analysis, ELM mathematics). The explanation for 'missingness' is the inability to link allocation to pupil data in some schools. One small consolation for the loss of data is that the loss of pupil data from whole schools means that the loss is distributed between treatment and control classes equally (since schools had equal numbers of classes in each allocation). As such, this should lessen the threat to internal validity arising from missing data.

For the disengagement outcome data, the final analysis sample was 1,572 pupils (61.3% of 2,564). This higher level of attrition arose because in addition to not being able to link allocations to pupils, the disengagement survey was not collected in some schools, and in some cases the responses could not be linked because there was no pupil name or UPN on the form.

Pupil and school characteristics

Table 6 sets out the pupil-level balance assessment. As the allocation is missing in places, this only includes data for those pupils for which the allocation was known. There was no missing data for year group or pupil sex. It is clear that year groups are imbalanced—with two-thirds of the treatment group being Year 4—but this is a difference that likely arose because we did not stratify by year group during allocation, hence we included year group in all of our analyses. Free school meal eligibility (everFSM) is perfectly balanced, as is gender. Similarly, at first glance the KS1 baseline measure looks imbalanced—there is nearly a ten percentage point difference between treatment and control groups. However, as described earlier, this baseline measure is a composite of two different testing regimes: the older 'levels' and newer categorical classification (below/at/exceeding 'standard'). Splitting this measure by year group—the final two rows in Table 6—illustrates that these measures are indeed balanced between groups.

Table 6: Baseline comparison—pupil data

	TREATMENT		CONTROL	
Pupil-level (categorical)	n/N (missing)	%	n/N (missing)	%
Year group: Year 4	753/1,103 (0)	68.3%	329/1,030 (0)	31.9%
Year group: Year 5	350/1,103 (0)	31.7%	701/1,030 (0)	68.1%
Ever FSM eligible	328/1,103 (32)	29.7%	307/1,030 (38)	29.8%
Pupil sex: Female	562/1,103 (0)	51.0%	517/1,030 (0)	49.0%
KS1 baseline: at/ above standard	837/1,103 (24)	75.9%	870/1,030 (29)	84.5%
KS1 Year 4: at/ above standard	529/753 (20)	70.3%	230/329 (7)	69.9%
KS1 Year 5: at/ above standard	308/350 (4)	88.0%	640/701 (22)	91.3%

In terms of school characteristics, the criteria schools needed to meet to take part in the trial was described in section ‘Participant selection’ (page 12). Finally, in the study SAP we proposed balanced checks on teacher characteristics, but because of the lack of data on teacher experience, age, and gender in many schools (or no data at all), this was not possible.

Outcomes and analysis

Figure 19 and Figure 20 in Appendix I show the distributions of the two outcome variables—the ELM raw score and the summed Engagement vs Disaffection with Learning questionnaire (EvDwLq) (Skinner et al., 2008). For both measures, we decided that they were suitable for linear modelling on the basis that both are approximately normally distributed. Note that the disengagement measure refers to a negative effect—higher scores mean *greater* disaffection/disengagement—so we would wish for these to be lower in the treatment group for evidence of effectiveness to be demonstrated. Table 7 shows the results of the primary analysis for the primary and secondary outcome measures and Table 8 presents the effect size estimation.

Table 7: Primary analysis

	Raw means						
Outcome	n intervention (missing)	Mean (95% CI)	n control (missing)	Mean (95% CI)	n in model (intervention; control)	Unadjusted differences in means	Adjusted differences in means
ELM maths	1,103 (*)	18.4 (18.0, 18.7)	1,030 (*)	19.7 (19.3, 20.1)	2,133 (1,103; 1,030)	1.35	0.02
Disengagement EvDwLq	885 (*)	65.9 (64.9, 66.7)	687 (*)	67.8 (66.9, 68.8)	1,572 (885; 687)	-1.96	-1.18

Table 8: Effect size estimation

	Intervention group		Control group				
Outcome	N (missing)	Variance of outcome	N (missing)	Variance of outcome	Pooled variance	Hedges g (95% CI)	p-value
ELM maths	1,103 (*)	42.03	1,030 (*)	41.05	41.99	0.04 (-0.17, 0.09)	0.957
Disengagement EvDwLq	886 (*)	183.32	692 (*)	164.12	175.67	-0.09 (-0.22, 0.04)	0.17

Intention-to-treat main effects

Reflecting the null result reported above, the *coefficient* for the treatment/control contrast for ELM mathematics is b -0.29 (se 0.43; 95% CI -1.13, 0.56). The full ITT model results for the ELM mathematics test and disengagement survey are given in Appendix I (Table 14 and Table 15).

FSM subgroup analyses

We pre-specified that we would look at subgroup effects for FSM-eligible pupils—acknowledging in both our protocol and SAP that this analysis would be under-powered (Sim et al., 2018; Sutherland and van Belle, 2018). To examine FSM subgroup effects, we ran the models specified in our SAP, but adding an interaction for FSM-ever eligibility and the treatment indicator (a cross-level interaction in a multilevel model). For both ELM mathematics and disengagement the results reflected the ITT analyses above: there were no differences between control and treatment FSM pupils on either measure. For the ELM maths test the result was *b* 0.29 (95% CI -1.53, 2.11). For the disengagement measure the result was *b* -2.127 (95% CI -7.00, 2.75) for n = 1,518 pupils in the final model, fewer than the main effects model, owing to the combined missing data for KS1 baseline, treatment allocation status, and FSM status.

Robustness and specification checks

We undertook one specification check where we re-ran the main ITT models as single-level (pupil-level) models with n-1 school dummy variables and cluster-robust standard errors for class membership. This did not change the substantive conclusion of 'no difference' for the ELMs maths measure (*b* -0.11; se 0.26; p .682; 95% CI -0.63, 0.41). For the disengagement measure, the results were consistent with the ITT model: pupils in the treatment group reported lower levels of disaffection/disengagement and in this model the difference was 'statistically significant' (*b* -1.28; se 0.58; p 0.03; 95% CI -2.43, -0.13).[11] Full model results are reported in Appendix I.

[11] Re-running this analysis using bootstrap re-sampling, which was not pre-specified, does not ostensibly change the point estimate but does move the p-value to *p* 0.079. This is not to say the result above is invalidated, but that we should be cautious about placing too much emphasis on the secondary outcome result, particularly in the context of the other problems the project faced.

Missing data analysis

In our SAP we said that we would predict missingness using the treatment indicator as a predictor. But with allocation missing, as it is here, then this strategy would not make sense. Instead, as agreed with the EEF, we look at range of results obtained by imputing all pupils missing treatment allocation (n = 431) to *control*, then doing the same but imputing to *treatment*. We acknowledge this is a sub-optimal approach because doing so means that half of the allocations in a given school would be wrong each time (that is, the true-treatment pupils would be correctly allocated in one model—the 'all to treatment' model—but incorrectly categorised in the 'all to control' model, and vice versa for the true-control pupils). The only other approach we could think of would be to iteratively allocate pupils to treatment and control, but that would be just as arbitrary and likely much more open to unknown/unknowable bias. The results for this analysis for the ELM maths outcome were that:

(1) When all pupils missing allocation were categorised as 'control', the treatment coefficient was b -0.29 with a p value of 0.51, with 95% CIs of -1.13, 0.56.
(2) When all were allocated to 'treatment', the result was b -0.33 (se 0.41; p 0.43; 95% CI -1.13, 0.48).

Repeating the exercise with the disengagement survey, the results were as follows:

(3) All pupils missing allocation to control: b -1.20 (se 0.81; p 0.14; 95% CI -2.79, 0.40).
(4) All pupils missing allocation to treatment: b -1.01 (se 0.81; p 0.21; 95% CI -2.59, 0.57).

Notwithstanding the caveats already set out, the results for the disengagement measure are consistent across specifications and show that those in the treatment group reported lower levels of disaffection/disengagement than those in the control group. This was reported in the ITT analysis, where the coefficient was b -1.18, which translates into a Hedges' g of -0.09 (-0.22, 0.04), as reported in Table 8 above. As such, it may be that there could be gains in terms of pupil engagement, but the study was not well powered to differentiate these from chance. However, that these gains, consistent with the programme theory as they may be, were not mirrored by improvements in mathematics ability appears to raise questions about the programme theory and the putative link between disengagement and learning. But, it is important to point out that the measure of disengagement is very general in nature—it is much more about overall disengagement rather than specifically related to maths. As such, the measure of disengagement used may have been the 'least worst' choice, but it still may not have been suitably focused on programme mechanisms to really challenge the engagement-learning-attainment pathway proposed.

Compliance analysis

Given the difficulties and problems with the data noted above, we were not able to undertake compliance analyses. This is a deviation from the SAP.

Cost

Box 2: Cost—key findings

- The per pupil per year marginal cost of implementing Digital Feedback over three years is estimated to be £14.51. The cumulative per pupil marginal cost over three years is £43.52. The marginal cost calculation of the intervention includes costs for the Explain Everything app, sending one teacher and RL to the initial training, and the cost of sending a teacher to the PLC meetings. The main cost to be covered by schools is for staff time to attend training sessions.
- Survey data suggests that the time it takes to set up Digital Feedback is unlikely to be burdensome for most teachers, but that it does require some consideration for extra time, particularly during the start-up phase
- More than half of Research Lead respondents and over two-fifths of teacher respondents reported *not* spending any unpaid time on the weekly running of Digital Feedback, suggesting that Digital Feedback was similar in terms of time demands to usual practice.
- Survey responses on costs incurred to set-up and implement Digital Feedback suggest that at the school level, the time it takes to set-up and implement Digital Feedback is unlikely to be an additional cost burden.

It should be noted that the findings presented are informed by responses in the midline and endline surveys, and therefore rely on self-report from respondents who may not have been representative of all staff. Taking these limitations into account, further details on staff time, direct staff costs, extension of staff contract hours, and supply cover, are presented below. In addition to this, the implementation team provided cost estimates for the implementation of the intervention. We provide cost estimates from an implementation team's perspective at the end of this section.

Marginal cost

The Digital Feedback in Primary Maths intervention was implemented over one school year (2017/2018) in 34 schools. For the estimated average marginal cost per pupil, we included the cost of the Digital Feedback application (£40 per class), the cost of sending one teacher and one Research Lead to the initial training (£250 in total for one day [teachers] plus half a day [Research Leads]), and of sending one teacher to the PLC sessions throughout the year. These costs were provided by the KYRA Teaching School Alliance. The Alliance noted that if a school sent an additional teacher, it either offered a free place or a half-price discount to the school; no costs were provided relating to schools sending RLs to the two additional half-day training sessions.

While some headteachers reported equipment costs and financial costs of hiring supply cover for the set-up and ongoing implementation of Digital Feedback, these are not included as marginal cost calculations in Table 9 and Table 10 (please see 'Staff costs' section below for details). Table 10 presents both the per pupil cost of using Digital Feedback over three years, and the cumulative per pupil cost over three years. For completeness, we also present the prerequisite costs and staff time estimates (in number of hours) in Table 11 since these are elements that should be taken into account by schools considering using Digital Feedback.

Cost per pupil estimate is based on 37.7 pupils per school. The average number of pupils per school was calculated by dividing the number of pupils in the intervention group (1,282) by the number of classes in the intervention arm (54 classes; 1,282/54=23.7). On average, intervention schools had 1.59 classes implementing Digital Feedback (54 intervention classes divided by 34 schools). Therefore, to estimate the per pupil cost over three years, the three-years-per-school cost was divided by 37.7 pupils (23.7 pupils x 1.59 intervention classes per school).

Table 9: Estimated cost per school for delivering Digital Feedback

Item	Type of cost	Annual cost	Total cost over 3 years	Total cost per pupil per year over 3 years
Digital Feedback application	Running cost per school	(£40 x 1.59 classes in each school) = £63.60	£190.80	
One-off training cost for teacher and Research Lead	Start-up cost per school	£250.00	£250.00	
PLC meetings costs for one year	Running cost per school	£400.00	(£400 x 3) = £1,200	
Total			£1,640.80	(£1,640.80/3/37.7) = £14.51

Table 10: Estimated average cost per pupil over three years of delivering Digital Feedback

	Year 1	Year 2	Year 3
Approximate cost per pupil per year	£18.93 (£63.60/37.7 + £250/37.7 + £400/37.7)	£12.30 (£63.60/37.7 + £400/37.7)	£12.30 (+ £63.60/37.7 + £400/37.7)
Cumulative cost per pupil per year	£18.93	£31.23	£43.52

Table 11: Summary of prerequisite costs and staff time estimates

Item	Type of cost	Annual cost/time	Total cost/time over 3 years
Prerequisite costs			
Equipment costs	Set-up cost per school	£2,228	£2,228
Staff time			
Initial staff training for teachers	Start-up cost per school	7hrs of teacher's time	7hrs of teacher's time
Initial staff training for Research Leads	Start-up cost per school	3hrs of SLT staff time	3hrs of SLT staff time
Professional Learning Community	Running cost per school	(8 x 1.5hrs) = 12hrs of teacher's time	(3 x 12hrs) = 36 hrs of teacher's time
Additional RL training	Running cost per school	(2 x 3hrs) = 6hrs of SLT staff time	(3 x 6hrs) = 18hrs of SLT staff time

Costs not included under marginal cost

We excluded from the cost-per-pupil calculation the cost of equipment since this was considered a prerequisite rather than a marginal cost and was seldom reported by survey respondents. Specifically, only five headteacher respondents gave this information across the midline and endline surveys (three out of eight in the former; two out of seven in the latter; the surveys were sent to the 34 headteachers from participating schools). Across these five schools, costs included new headphones at £200, upgrading of tablets at £100, and new tablets ranging from £300 to £8,000, with an average of £2,228 (range £100–£8000, median £300). One headteacher reported not being sure about what costs their school had incurred to have all the required resources in place to start with Digital Feedback, while another headteacher reported purchasing applications at the cost of £60, though it is not clear what app the respondent was referring to. It should be noted that, in this trial, many schools paid for the app themselves, despite being offered a reimbursement to cover these costs.

Staff time and cost

As per the EEF's guidance on cost evaluation (EEF, 2016), we excluded school staff time from the cost per pupil per year estimate but present our findings on these in the narrative below. EEF guidance suggests that instances where the schools do not incur a financial cost, but provide cover by re-allocating existing resources, should be reported in terms of time (number of hours) rather than in pounds.

Based on responses from schools, the start-up and the running of the intervention required the most amount of unpaid staff time. However, it should be noted that the amounts of time reported ranged widely across participants, and many reported not spending any extra time delivering Digital Feedback. This is discussed in further detail below.

The main cost to schools was to absorb the cost of covering teachers' and Research Leads' time to attend trainings and PLC meetings. Training included an initial full day of training for teachers (seven hours, 09:00–16:00) and half a day of training for RLs who were senior staff (members of SLT; three hours, 13:00–16:00). In addition to the initial training, further training was provided in the form of PLC meetings for teachers (eight sessions of, on average, 1.5 hours, excluding travel time),[12] and two half day sessions for RLs (three hours each, 13:00–16:00).

Use of unpaid staff time to deliver Digital Feedback

In the midline and endline surveys, we asked teachers, Research Leads, and headteachers for their estimates of the amount of unpaid time they spent (or in the case of headteachers, how much time their staff spent) in the start-up phase and ongoing implementation of Digital Feedback.

Overall, more than half of Research Lead respondents and over two-fifths of teacher respondents reported not spending any unpaid time on the weekly running of Digital Feedback, suggesting that Digital Feedback was similar in terms of time demands to usual practice.

Despite this, it should be noted that results from the survey questions asking about unpaid extra time spent on Digital Feedback at both start-up and implementation (see Table 12) suggest that unpaid digital time differed for different participants. For example, on average teachers and Research Leads reported spending an extra hour of unpaid staff time on implementing Digital Feedback, but the range varied from between one to three hours. This suggests that the time it takes to set up and implement Digital Feedback is unlikely to be burdensome for most teachers, but that it does require some extra time, particularly during the start-up phase.

Table 12: Average unpaid extra time spent by staff on Digital Feedback in the start-up phase and during the weekly implementation

	Average unpaid time in start-up phase	Average unpaid time for implementation
Teachers (n = 42)	2.2 hours (range: 0–10 hours)	1.01 hours (range: 0–3 hours)
Research Leads (n = 33)	0.95 hours (range: 0–5 hours)	0.53 hours (range: 0–5 hours)
Headteachers (n = 14)	3.21 (range: 0–10 hours)	0.93 (range: 0–3 hours)

Of the participants that did report spending unpaid time on delivering Digital Feedback, only three teachers chose to give details about the specific types of activities they did during unpaid time and all were related to the digital feedback aspects of the intervention, such as making videos or reporting digitally.

Staff costs

We asked headteachers to provide details of any staff costs incurred during the start-up and implementation phases of Digital Feedback in both the midline and endline surveys. We saw these staff costs as potentially falling into two categories: supply cover and the need to buy in additional staff, either by hiring new staff or extending the contract hours of existing staff.

The majority of respondents reported that they had not hired supply cover. Where respondents (n = 9) provided numerical estimates for the cost of hiring supply cover, estimates ranged from £200 to £2,000. The average of the reported supply cover costs was £700. Please note that this estimate encompasses supply cover costs for the start-up and ongoing implementation of the intervention. Only four headteachers reported hiring supply staff for the ongoing implementation of Digital Feedback (range £400–£800, average £550). None of the headteachers reported hiring new staff or extending staff contract hours for Digital Feedback.

[12] The majority of PLC meetings the evaluation team observed were scheduled for one hour, while only one was scheduled for two hours. Based on a conversation with a PLC coordinator during a PLC observation, it was noted that some teachers had to travel long distances to attend the meetings. We did not obtain an average travel time to be able to incorporate this into the cost estimate. Similarly, this was the case for some, but not all, teachers; in some instances the training took place in the school where some teachers were based, meaning that for those teachers no travel costs were incurred.

Taken together, this suggests that, at the school level, the time it takes to set up and implement Digital Feedback is unlikely to be an additional cost burden.

Costs incurred by implementation team

As reported by the implementation team, the overall expenditure for the trial year amounted to £148,947.33, 7% of which related to the training of Research Leads and teachers and 34% of which related to implementation and project management costs. The PLC facilitation (and associated travel costs) amounted to approximately 38% of the budget. Resource development and training materials was approximately 19% of the budget.

Implementation and process evaluation

The implementation and process evaluation (IPE) explored the perspectives of teachers, Research Leads, and headteachers on the training and resources provided for Digital Feedback, as well as the implementation and perceived impact of the programme. As previously mentioned in the methods section, the IPE sought to answer the following questions:

- Was the intervention implemented with fidelity in the intervention schools?
- What factors and initial conditions appear to explain variation in fidelity of implementation?
- What appear to be the necessary conditions for success of the intervention?
- What were the barriers to delivery?

The data gathered for the process evaluation is summarised in Table 13.

Table 13: Overview of data gathered as part of the implementation and process evaluation

Data type	Intervention phase	Participants (sample size)	Collection timeline	Reference code in text
Baseline survey	Pre-intervention	Teachers (n = 32 intervention; n = 23 control) Research Leads (n = 18) Headteachers (n = 19)	September 2017	SUR [teacher/RL/head teacher] baseline
Midline survey	Intervention	Teachers (n = 23 intervention) Research Leads (n = 16) Head teachers (n = 8)	February 2018	SUR [teacher/RL/head teacher] midline
Endline survey	Post-intervention	Teachers (n = 19 intervention; n = 12 control) Research Leads (n = 17) Head teachers (n = 7)	June–July 2018	SUR [teacher/RL/head teacher] endline
Observation of training sessions	Pre-intervention	Teacher training (n = 3) Research Lead training (n = 2)	June 2017	N/A
Observation of PLC sessions	Intervention	n = 3	May–June 2018	PLC1, PLC2, PLC3
Focus group*	Intervention	With teachers (n = 1, six participating teachers)	May 2018	FG
Analysis of PLC session RAG ratings**	Intervention	RAG ratings cover all PLC clusters (n = 5; Nottingham, Lincoln A, Lincoln B, Rotherham/Doncaster/Scunthorpe A, Rotherham/Doncaster/Scunthorpe B) and all 8 PLC sessions	June 2018	RAG ratings
Case studies** - **Analysis of coaching records**	Intervention	From 3 case study schools	February–September 2018	N/A
- **Analysis of feedback videos**	Intervention	Obtained from 2 case study schools; analysed random sample of 10 feedback videos	February–September 2018	N/A
- **Teacher interview**	Post-intervention	n = 1	November 2018	INT

* This was intended to be an observation of a PLC session; however, as the PLC coordinator was unable to attend, the researcher present took the opportunity to collect qualitative data from teachers via an impromptu focus group.
**While the analysis of these materials took place post-intervention, materials obtained covered the span of the intervention (for example, coaching records covered all meetings teachers had with Research Leads during the intervention).

All activities listed above, with the exception of the baseline survey, took place post-randomisation. Randomisation took place in May 2017. The intended samples for survey completion were the following: 28 headteachers, 28 Research Leads, 41 treatment teachers, and 41 control teachers. However, please note that these numbers include participants with dual roles, that is, five headteachers were also assigned as RLs, seven RLs were also assigned as control teachers, and one RL was also assigned as treatment teacher.

It should be highlighted here that the findings reported in this section draw largely on survey, focus group, and interview findings, which represent respondents' self-reports and perceptions and may not be representative of the sample as a whole. We also have to acknowledge that the surveys are not random samples, hence selection bias may influence responses. For example, those most motivated about Digital Feedback may have been those most willing to respond

to the survey. Furthermore, the surveys did not track respondents across time-points and therefore should not be interpreted as longitudinal. Finally, as can be seen in the number of survey respondents reported above, response rates were low and generally decreased during the course of the evaluation. Nevertheless, the purpose of the IPE was to gain a better understanding of how the intervention was implemented in practice and participant perceptions of the approach. As such, these activities allowed us to generate some insights into why/how Digital Feedback may/may not have impacted pupils' mathematics outcomes.

In the following sections, analyses of the different data collected for the IPE are presented. Please note that survey figures include only the 'strongly disagree', 'agree', and 'don't know' response categories to improve readability of the figures. Please refer to Appendix J where response counts for all response categories of the survey questions are provided.

Implementation and fidelity

Box 3: Implementation and fidelity—key findings

- Overall, IPE data suggests that there were low levels of fidelity in the implementation and daily running of the intervention during the trial.

Training

- Baseline survey responses showed that the majority of teacher respondents attended the one-day training session. Only slightly more than half of Research Lead respondents reported attending the half-day training.
- Teachers and Research Leads reported having positive views on the training (clarity, level of detail, possibility to ask questions, trainer responses to questions) and most reported feeling the training prepared them well. However, at midline most survey respondents reported feeling 'somewhat prepared' (rather than 'very prepared') to implement Digital Feedback, with some Research Leads respondents reporting not feeling prepared at all.
- Observations of teacher and Research Lead training sessions indicated that there was some confusion by participants on what the intervention would be about and what the goals of the training were. These observations also highlighted that some details about the training content (implementation practicalities) could have been strengthened. Nonetheless, some participants expressed high levels of enthusiasm to participate in the trial and contribute to research.

Provision of immediate and delayed feedback

- Intervention survey responses showed a general decrease in the amount of digital feedback (immediate and delayed) given by teachers as the trial progressed.
- Potential factors hindering the provision of digital feedback identified by participants were (1) lack of equipment and/or the cost of equipment, (2) difficulty in using, and cost associated with, the Digital Feedback application, and (3) the time taken to provide video feedback.
- Survey respondents and focus group participants noted that embedding the use of Digital Feedback in daily practice was difficult (for example, the use of the Explain Everything app or of iPads to record feedback was cumbersome and time consuming since in some schools individual pupil log-ins were needed). Further guidance on how to better to embed the use of the app would have been helpful at training stage.
- Some participants reported making adaptations to the intervention by, for example, allowing pupils to record feedback videos or by recording generic 'explanation' videos before lessons that could be provided to various pupils.

Coaching and PLC sessions

- At the beginning of the intervention, teachers, Research Leads, and headteachers reported seeing the coaching role of Research Leads as important for the success of the intervention.
- In practice, process evaluation data (survey, PLC observations, and review of coaching records) indicates that coaching meetings were irregular, with some teachers reporting not having met with their Research Lead at all. The endline survey showed a reduction in reported frequency of meetings between Research Leads and teachers compared to the midline survey.
- A potential factor hindering coaching sessions taking place is the busy schedule of Research Leads who tended to have senior leadership roles in their schools.
- Teachers receiving coaching found meetings with their Research Lead useful in general (midline and endline surveys).
- Survey responses indicate that teachers attended PLC meetings, but a decrease in frequency was reported at endline compared to midline. At endline, less than half of teacher survey respondents

reported joining PLC meetings. The review of PLC Red-Amber-Green (RAG) ratings corroborates this and showed significant variation in PLC attendance throughout the trial.
- Most teachers that attended PLC meetings reported finding PLC meetings at least 'somewhat useful' or 'very useful'. Views that these meetings were 'very useful' remained relatively constant between midline and endline (slight increase observed at endline) but the smaller sample size should be kept in mind.
- During PLC meeting observations teachers were engaged in discussions and explicitly noted how much they valued the sessions.

Use of diagnostic assessments
- In the midline and endline surveys the majority of survey respondents reported that diagnostic assessments were used at least once a week by teachers.
- There were variations in the reports of teachers, Research Leads, and headteachers on how diagnostic assessments were used by teachers (for example, to focus on teaching areas that the whole class struggles with or to track progress of a class over course of the year).
- The majority of teachers reported (at endline) using diagnostic assessments to focus teaching on areas that the class seemed to be struggling with (above 80%) and to track progress made by the class as a whole over the year (above 50%).
- Reports on the use of diagnostic assessments may suggest that during the course of the intervention participants used Digital Feedback less as a tailoring or targeting tool and more as a global classroom tool.

This section discusses the implementation and fidelity of Digital Feedback in Primary Maths with respect to its four key factors as identified in the logic model—high quality feedback, the use of digital feedback, PLCs, and coaching. We also discuss the perceived effectiveness of the training delivered to school staff implementing the intervention.

Participant attendance at training and their perceptions

Before the start of the intervention, teachers had a one-day training session available to them and Research Leads a half-day session. Participants were asked about their experience of the training in the midline survey. Survey responses show that the majority of intervention teachers in the midline survey (87.0%; 20 out of 23) reported that they had attended the one-day training session. In general, teachers were positive about the training with respondents reporting that information was detailed and useful, that they had been able to ask questions specific to their class or school, and that the trainer provided useful answers to questions (see Figure 4). (However, please note that teachers responding to the surveys may have been the most positive pool of teachers, and therefore willing to engage with the evaluation.)

In contrast, only slightly more than half of Research Lead respondents in the midline survey (56.3%; 9 out of 16) reported that they had attended the half-day training session. Research Leads' half-day training session also included information on coaching. When attending a Research Lead training session, we observed that the training focused less on coaching than on other aspects of the intervention, such as effective feedback and the use of the Digital Feedback application. Perhaps as a consequence of this, Research Leads who reported attending the training session provided more mixed feedback on the effectiveness of the training for coaching—as can be seen in Figure 5 below—compared to feedback on the tablet and application (shown in Figure 4). The combination of initial training possibly not sufficiently covering the 'coaching' role and the mixed feedback received on the relevance of this training raises the possibility that Research Leads may not have been fully prepared to fulfil their coaching role.

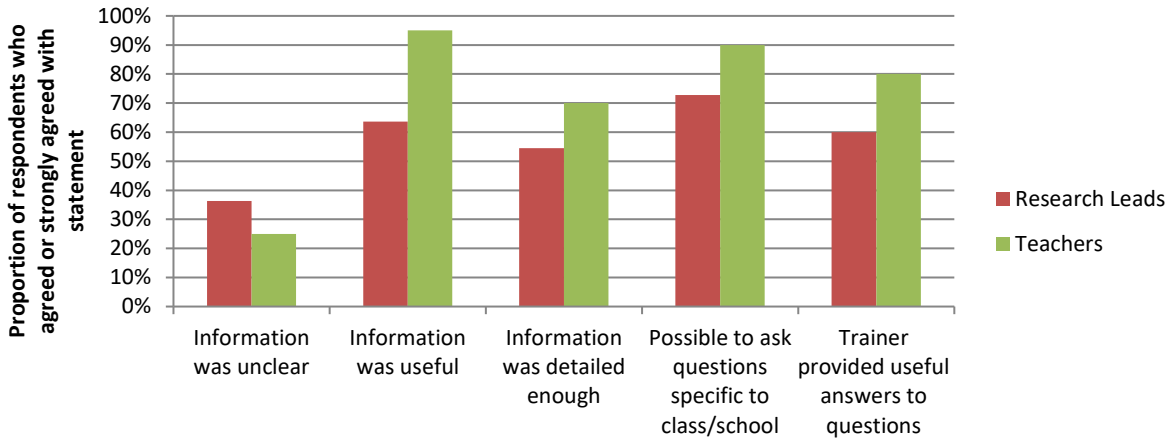
Given the coaching and support role that Research Leads were meant to play, this raises concerns that a key element of the intervention may not have been as prepared as necessary.

Furthermore, despite positive reports on the training in the surveys, overall impressions of the teacher and Research Lead training observations hint that further information about the training could have been provided to participants prior to the training, as well as that some of the detail in the content of the training around implementation practicalities could have been strengthened. During training observations we noted that a considerable proportion of participants did not have a clear understanding of what the training would be about (its main goal), or what they would be committing to prior to joining the session.

Moreover, we observed in teacher and Research Lead training sessions that the digital element of the intervention received less attention than other aspects, such as elements of good or effective feedback. Even when the application

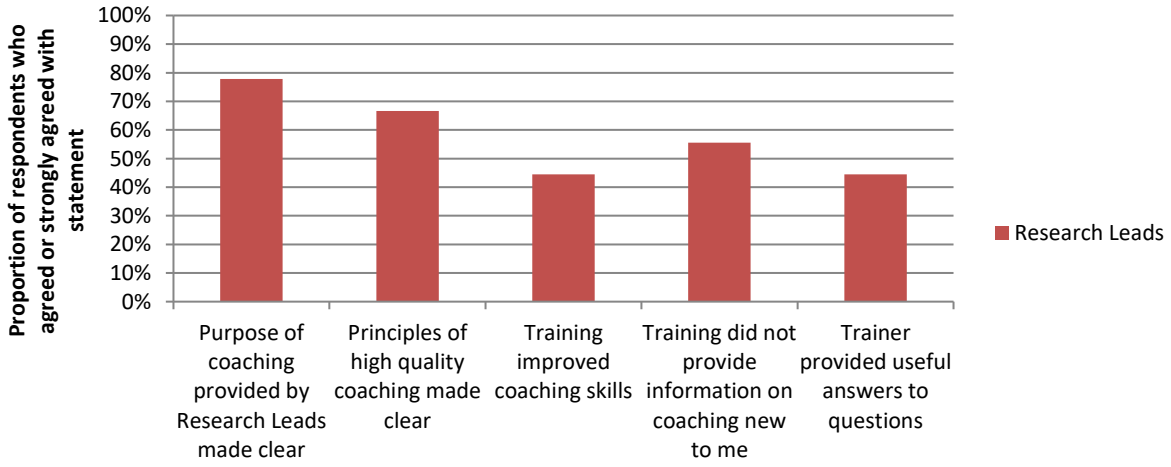
was discussed, we observed that the focus was on the practical elements—such as how to use the application and how to save videos—rather than perhaps more substantive issues, such as how to use the application effectively (OBS RL training). During the training observations we also noted that some implementation details were not discussed and remained unclear to participants (for example, how to implement the intervention when a year group has streamed classes or mixed age groups, or how to address issues around video storage space and Wi-Fi connections). Furthermore, in our PLC observations towards the end of intervention implementation, some teachers reported that they felt too much information had been covered in the initial training session (PLC3). These points suggest that further work might have been needed prior to training delivery to think about how to communicate the intervention to potential participants and to define the content of the training (level of detail and amount). Nonetheless, some participants, in particular Research Leads, expressed being keen and excited to participate in the trial and to contribute to research; teachers remained somewhat more sceptical (OBS teacher training, OBS RL training).

Figure 4: Research Lead and teacher survey respondents' feedback on the training on the use of the tablet and the Digital Feedback application (midline)



Teacher, n = 20; Research Lead, n = 11.

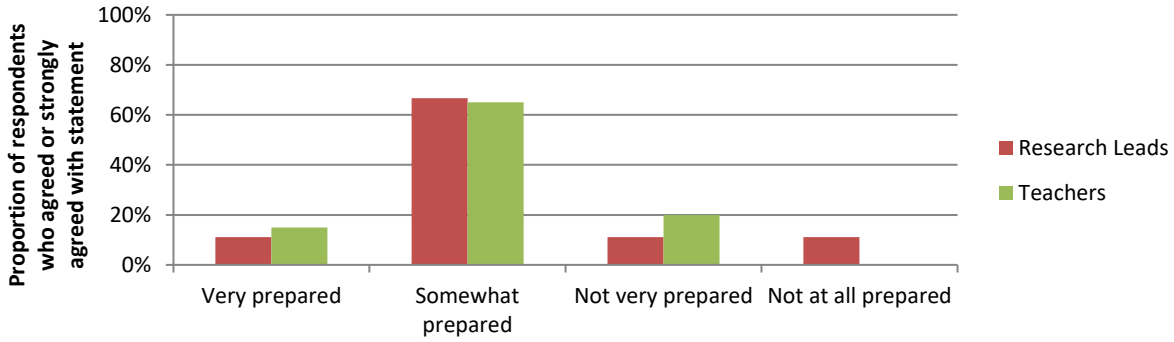
Figure 5: Research Lead respondents' feedback on the training on coaching (midline)



Research Lead, n = 9.

When Research Leads and teachers were asked in the midline survey the extent to which they felt the training had prepared them to support teachers using Digital Feedback in their school and to use Digital Feedback in their class, respectively, the majority of respondents reported feeling 'somewhat prepared' (Figure 6).

Figure 6: Research Lead and teacher survey respondents' response on their preparedness following training (midline)

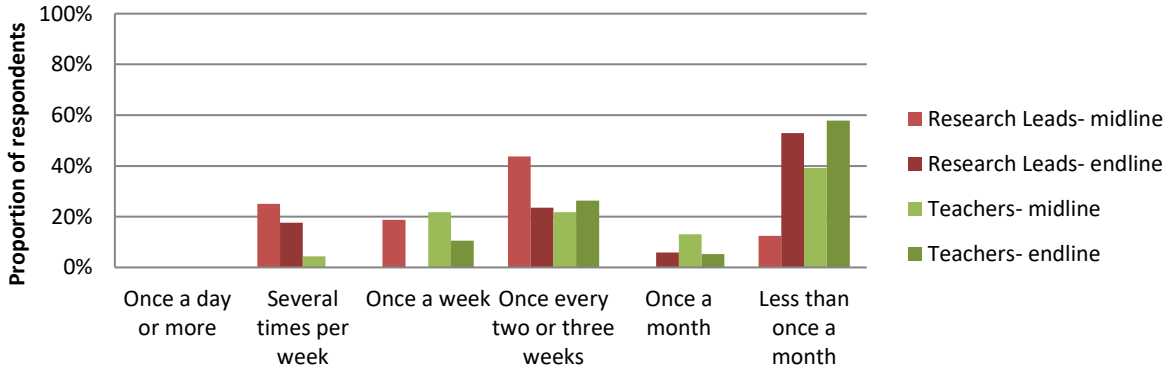


Teacher, n = 20; Research Lead, n = 9.

Frequency of providing immediate and delayed digital feedback via the Digital Feedback application

The intervention aims to improve teacher practice and pupil learning by changing the type and way feedback is provided. There were two anticipated types of digital feedback teachers would use with pupils—immediate or delayed—though the frequency with which feedback was given was at the discretion of teachers. Figure 7 and Figure 8 show teachers' and Research Leads' responses in the midline and endline surveys on how often they used immediate and delayed digital feedback respectively. It appears, from the responses available, that digital feedback was sometimes not used or given very infrequently (Figure 7). For example, very few teachers reported giving digital feedback 'several times per week' (midline 4%, 1 out of 23; endline none out of 19). This is at odds with the fact that a full-time teacher would be teaching maths to the same class each week and would have therefore had the opportunity to provide Digital Feedback more regularly. However, we did not have a question on the overall frequency of other types of feedback for comparison. Research Leads reported higher frequencies of immediate feedback overall ('several times per week' midline: 25%, 4 out of 16; endline: 18%, 3 out of 17), but that may be because their role would explicitly lead to greater exposure to this behaviour (or the teachers they were in contact with regularly were the most enthusiastic about using Digital Feedback). Overall, 'less than once a month' had the highest number of responses (midline: Research Leads, 12%, 2 out of 16; teachers, 39%, 9 out of 23; endline: Research Leads, 53%, 9 out of 17; teachers, 58%, 11 out of 19).

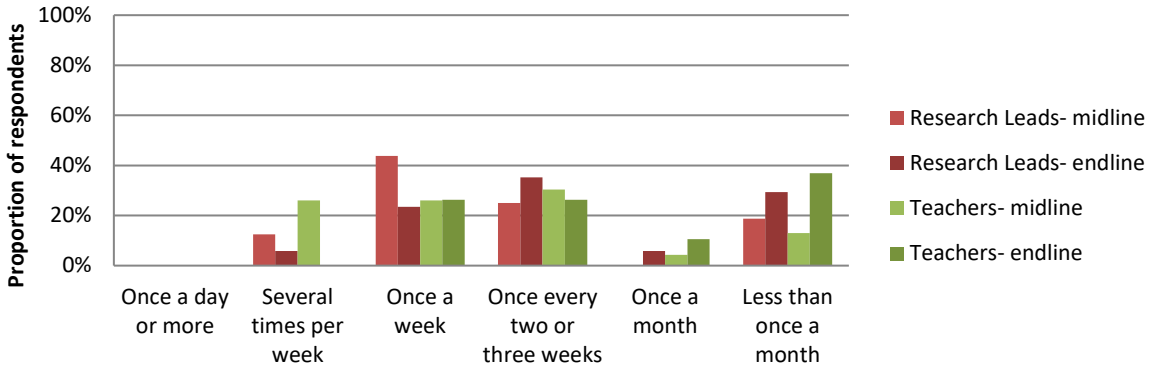
Figure 7: Research Lead and teacher survey respondents' response in the midline and endline surveys on the frequency of provision of immediate digital feedback



Midline: teacher, n = 23, Research Lead, n = 16; endline: teacher, n = 19, Research Lead, n = 17.

Similar to immediate feedback, there were mixed responses from Research Leads and teachers about how often delayed digital feedback was used, and none of the Research Lead and teacher respondents stated that delayed digital feedback was provided once a day or more. In the midline survey, Research Leads were most likely to report that delayed digital feedback was given once a week (43%, 7 out of 16), while teachers were most likely to report that it was given once every two or three weeks (30%, 7 out of 23) as shown in Figure 8. In the endline survey, this pattern had shifted; the most common response from Research Leads was once every two or three weeks (35%, 6 out of 17), while the most common response from teachers was less than once a month (37%, 7 out of 19). Again, with both figures we have to caution that there were few responses overall so the results are indicative rather than definitive.

Figure 8: Research Lead and teacher survey respondents' response in the midline and endline surveys on the frequency of provision of delayed digital feedback



Midline: teacher, n = 23, Research Lead, n = 16; endline: teacher, n = 19, Research Lead, n = 17.

Teachers reported that cost, difficulty, and time were reasons for not using the app

Data collected revealed a number of potential reasons why teachers may not have been using the app to provide digital feedback to pupils: (1) a lack of equipment or the cost of equipment, (2) difficulty in using the Digital Feedback application and the cost associated with it, and (3) the time taken to provide video feedback. Each of these identified difficulties are described below in more detail.

Equipment cost

First, a lack of equipment or hardware may have accounted for why some teachers were not providing digital feedback. Informants in the PLC observation (PLC3), focus group, as well as the midline and endline surveys (including teachers, Research Leads, and headteachers), pointed to the lack of tablets in their schools or classrooms (and their associated cost) as a barrier to implementing Digital Feedback (see Box 4 for some examples).

Box 4: Barriers to providing digital feedback—open-ended survey answers

> *'It is […] not readily accessible if you do not have enough iPads for pupils' (SUR teacher, midline).*
>
> *'The lack of ICT equipment hinders the regular use of Digital Feedback' (SUR RL, midline).*
>
> *'The cost of providing tablets for digital feedback [was a concern about the intervention]' (SUR headteacher, endline).*

Ease of use of the Explain Everything app

Second, some informants pointed to problems associated with the Digital Feedback application. Some teachers felt that **using the application or iPads to record digital feedback was cumbersome** (PLC2). In the midline and endline surveys, teachers were asked the extent to which they agreed it was difficult to provide digital feedback using the app. Although only a minority of respondents agreed or strongly agreed with this statement in the midline survey (9 out of 23), this pattern reversed in the endline survey—13 out of 19 agreed with the statement—**suggesting that by the end of the implementation period, more teachers may have found it difficult to use the app than not**. During one observation of a PLC session, the PLC coordinator speculated that problems using the application may have dampened teachers' initial excitement about using it (PLC1). Another issue raised was the cost of the application itself (SUR teacher, midline; PLC2).[13] In response to these issues, some teachers reported that instead of using the app, teachers were simply recording digital feedback using the camera on their tablets and then sharing the videos with their pupils (PLC2; PLC3). Taking this alternative approach was suggested by the delivery team to schools that were slow to purchase the Explain Everything app.

[13] As part of this trial, the cost of the Explain Everything app was covered. However, the Delivery Team confirmed that not all participating schools reclaimed the costs to purchase the application. Therefore, some schools absorbed this cost themselves.

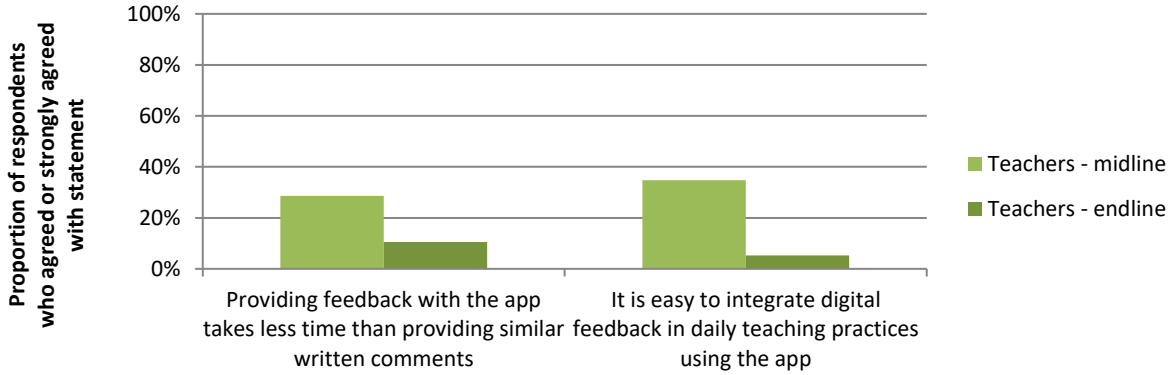
It is worth noting that as part of the trial, schools should have had the cost of the app covered—they needed to invoice the intervention team to get the cost reimbursed. However, the implementation team noted that some schools did not reclaim the costs for the app. Similarly, as part of the eligibility criteria, schools were expected to have enough iPads for pupils (roughly one tablet for every four to five pupils). The fact that some intervention participants identified equipment and app costs as barriers to implementation suggests that some of these requirements may not have been met by all schools during the trial.

Time limitations

Third, several teachers, Research Leads, and headteachers reported in the midline and endline surveys **that the amount of time taken to record video feedback was a barrier to implementation**. Teachers and Research Leads felt that using Digital Feedback had been 'time consuming' (SUR RL midline; SUR RL endline; SUR headteacher endline), potentially, in particular, 'with large classes' (SUR RL endline), and 'ha[d] not saved marking time' (SUR teacher midline). As can be seen in Figure 9, when asked if providing feedback with the app took less time than providing similar written comments, only a minority of teacher respondents in both the midline (26%, 6 out of 23) and endline (11%, 2 out of 19) surveys 'agreed' or 'strongly agreed'. Anecdotal evidence from the focus group and interview suggested that teachers may also have felt that it took more time to prepare the video feedback than give feedback to a pupil in person (FG, INT).

Relatedly, several teachers and Research Leads reported that it had been difficult to make Digital Feedback a part of the daily routine. When asked if it was easy to integrate digital feedback into daily teaching practice, only a minority of teacher respondents in both the midline (35%, 8 out of 23) and endline surveys (5%, 1 out of 19) 'agreed' or 'strongly agreed' with the statement (see Figure 9).

Figure 9: Teacher respondents' level of agreement with statements on using the app to provide digital feedback



Midline, n = 23; endline, n = 19.

The challenge of integrating Digital Feedback may have been particularly relevant to immediate feedback. In the midline and endline surveys, both teachers and Research Leads reported that embedding immediate feedback into the classroom had been challenging as shown in Box 5. However, caution is needed when interpreting these comments since they could be influence by a lack of practice or confidence in using the app or to how well (or not) the training prepared teachers in using the app.

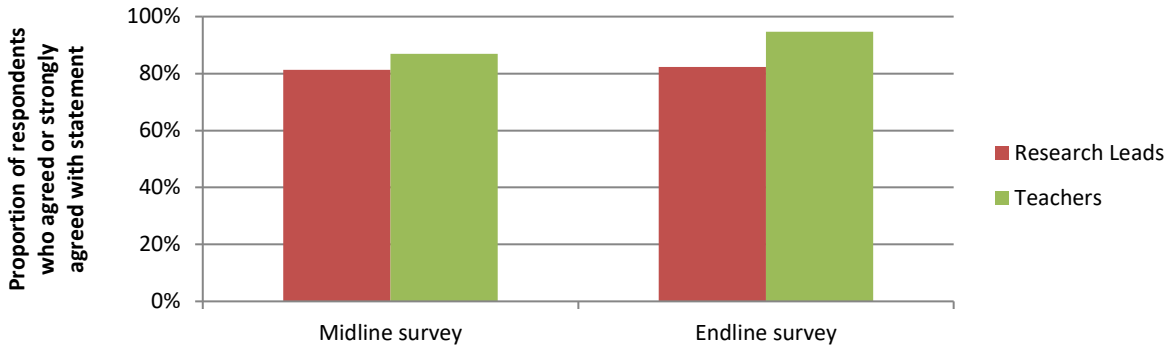
Box 5: Embedding Digital Feedback in daily practice—open-ended survey answers

> '*Although I think that digital feedback is a great way of providing detailed feedback to children, I really struggle to fit the feedback on iPads into my lessons*' (SUR teacher midline).

> *'It is not possible to provide [digital feedback] to children more than a couple of times in a lesson'* (SUR teacher midline).
>
> *'It is hard to explain why but it has been difficult to set up a routine that works with the immediate feedback'* (SUR teacher endline).
>
> *'The amount of time to make the feedback, when done in class […], can inhibit effective use for all pupils'* (SUR teacher endline).
>
> *'Setting up the routines and practices with immediate feedback has proved to be the most challenging part to get working in practice'* (SUR RL midline).
>
> *'I really believe it should work, but it has proven to be challenging to just get it going and embed it into practice'* (SUR RL endline).

When asked, in the midline and endline surveys, the extent to which they agreed that it was difficult to fit in immediate digital feedback during class, more than 80% of Research Lead respondents (midline 13 out of 16; endline 14 out of 17) and more than 85% of teacher respondents (midline 20 out of 23; endline 18 out of 19) agreed or strongly agreed with the statement (Figure 10).

Figure 10: Survey responses for the statement 'During a class it is difficult to fit in immediate digital feedback'



Midline: teacher, n = 23, Research Lead, n = 16; endline: teacher, n = 19, Research Lead, n = 17.

Teachers reported some adaptations to the intervention which were perhaps due to the perceived difficulties in providing video feedback. First, several teachers reported that they encourage pupils to record their own videos (PLC1, PLC3). For instance, in one of the PLC sessions observed by the evaluation team, a teacher presented an exercise in which they had asked a pupil to give 'step-by-step' feedback on an exercise completed by a fellow pupil (PLC1). Second, informants also reported recording videos that could be provided to more than one pupil, though this was not necessarily as effective as tailored feedback (Box 6).

Box 6: Digital Feedback intervention adaptations—open-ended survey answers

> *'I like the use of videos […] but I struggle to do these during lesson due to time. I often ask one of the children to film me teach[ing] an example question and will hand that out to children who are unsure' (SUR teacher midline).*
>
> *'A good use of digital feedback was a short video for a common misconception so that many could use it but only one needed to be made by the teacher' (SUR RL endline).*
>
> *'[I] have made videos for groups of children but these are less personal. Often, children still have questions and will still come to ask me' (SUR teacher endline).*

A teacher interviewee reported that they would prepare such videos, on certain methods or errors, prior to lessons so that they could share them with the pupils who needed more clarification (INT). In our review of a random sample of ten feedback videos, we found that four of these featured teachers explaining how to solve a problem (rather than focusing on feedback for specific pupils). As discussed in the previous sub-section, it may have been that during the course of the year, participants began to increasingly use Digital Feedback as a global classroom tool.

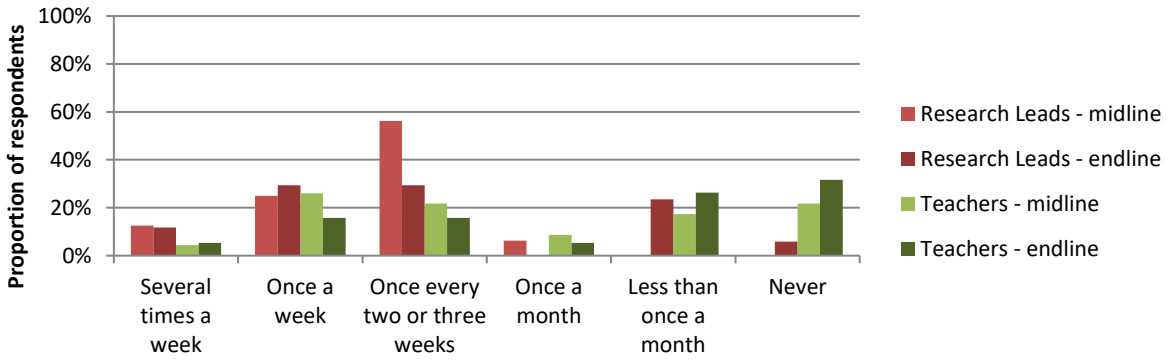
Coaching and PLC attendance

The role of the Research Lead was seen as important by headteachers. In the baseline survey, all respondents thought that the Research Lead could contribute to the success of Digital Feedback; in the endline survey, five of seven headteacher respondents (71.4%) reported feeling the same. Research Leads themselves reported feeling the same way: when asked in the baseline survey, almost all respondents (94.4%; 17 out of 18) thought that the role could contribute to the success of the intervention.

When it comes to the role of Research Leads, the intervention logic model outlines that Research Leads would conduct weekly class visits to observe teacher feedback practices. Looking at the survey responses related to the training (discussed above) it appears that Research Leads did not participate fully in the training and that the coaching aspect was perhaps not as developed at training as other elements of the programme.

As for coaching in practice, our analysis of coaching records from three case study schools showed that at least some of the Research Leads were conducting observations, though it was unclear how often these took place. The expectation was for Research Leads to conduct weekly class visits and provide feedback as part of coaching sessions. However, our analysis of coaching records found that coaching frequency varied, with sessions taking place three times a week in one school and once or twice a month in the other two schools. Survey respondents also provided mixed responses as to whether weekly coaching sessions were actually implemented, as can be seen in Figure 11.

Figure 11: Research Lead and teacher survey respondents' responses in the midline and endline surveys on how often they met



Midline: teacher, n = 23, Research Lead, n = 16; endline: eacher, n = 19, Research Lead, n = 17.

The majority of Research Lead respondents in the midline and endline surveys reported that meetings with teachers took place at least once every two or three weeks. While this was also true for teacher respondents, there was also a significant minority of teacher survey respondents (over 20%) who reported meeting less than once a month (midline 4 out of 22; endline 5 out 19), or even never (midline 5 out of 22; endline 6 out 19). Indeed, during our observations of PLC meetings there was mixed feedback from teachers on coaching. Where the discussion touched on elements of PLC coaching, it became evident that most teachers present had not had frequent coaching sessions with Research Leads, with some teachers reporting having never been coached (PLC2, PLC3, FG). For some of the teachers, coaching sessions took place more regularly at the beginning of the trial but frequency decreased as the school year progressed (PLC3).

One potential reason for the lack of regularity in coaching sessions is that Research Leads tended to be members of the senior leadership team or senior management within the school, which made it difficult to find time to meet (PLC2, PLC3). Some of the teachers expressed ambivalence about the necessity of the coaching sessions, stating they had 'done perfectly well without it' (PLC2). These findings suggest that further consideration should be given to who are selected as Research Leads to ensure that they are able to dedicate time to the intervention.

Participants who had experienced at least some coaching sessions (regardless of regularity or frequency) appeared to think that they went well. The majority of teacher participants in both the midline (77.8%, 14 out of 18) and endline surveys (61.5%, 8 out of 13) reported that they found the meetings with their Research Lead somewhat or very useful. Some teachers reported that the coaching sessions may have been more useful if Research Leads also had hands-on

experience with Digital Feedback or other elements of the intervention such as PLC meetings (PLC3, FG). For example, during the focus group, participants suggested that Research Leads and teachers would benefit from Research Leads attending PLC meetings since this could help increase Research Leads' understanding of the intervention and the support they provide to teachers.[14] This reinforces the point that Research Leads' training could be improved. More positively, almost all Research Lead respondents in the midline (93.8%, 15 out of 16) and endline surveys (94.1%, 16 out of 17) reported that the teachers they coach were somewhat open or very open to suggestions and advice.

According to the intervention logic model, teachers were expected to attend PLC meetings once a month throughout the duration of the intervention (eight PLC meetings in total). In the midline survey, 69.6% (16 out of 23) teacher respondents reported attending most of the meetings (4 or more meetings at that point), though this decreased to fewer than half (47.4%, 9 out of 19) of teacher respondents in the endline survey (who reported attending seven or more meetings). Our review of the PLC RAG ratings found that attendance from session to session varied greatly, from 15% to 100% (but it should also be noted that PLC RAG ratings provided to the evaluation team were not comprehensive and did not include all PLC meetings). PLC coordinators sometimes received apologies from teachers who were unable to attend meetings, citing, for instance, work or family commitments (PLC1). Another PLC coordinator noted that distance to the PLC meeting location may have been another reason for teachers' absences, though unexpectedly, it was the teachers who had to travel furthest who also had the highest attendance (PLC2).

The majority of teachers who reported attending the *PLC meetings* found them at least 'somewhat useful'. In the midline survey, 45.5% of respondents (10 out of 22) found the PLC meetings somewhat useful while 40.9% (9 out of 22) found them very useful. In the endline survey, 29.4% (5 out of 17) and 41.2% (7 out of 17) of respondents reported finding the PLC meetings somewhat and very useful respectively. However, there was an increase between the surveys in the proportion of participants reporting that these meetings were 'not very useful'. At endline, 29.4% of respondents (5 of 17) reported that the Digital Feedback training was 'not very useful' (none reported finding the training 'not useful at all'), while at midline this was 4.5% (1 out of 22) for 'not very useful' and 9% (2 out of 22) for 'not useful at all'. Nonetheless, during the evaluation team's observations of PLC meetings, we found that most teachers were very engaged in discussions and some explicitly noted how much they valued the meetings. Similarly, teacher informants in the focus group and interview (as shown by the quote below) were very positive about the PLC meetings.

> *'This is something within teaching I've noticed, that people don't want to sacrifice anything to make change, whereas last year we sacrificed one afternoon every month, and it really helped. It helped my practice, it helped my knowledge […] so that was one thing that was very, very positive, to do the PLCs […] to go every month, to have people who have the same problems, you could talk about problems' (INT).*

Use of diagnostic assessments

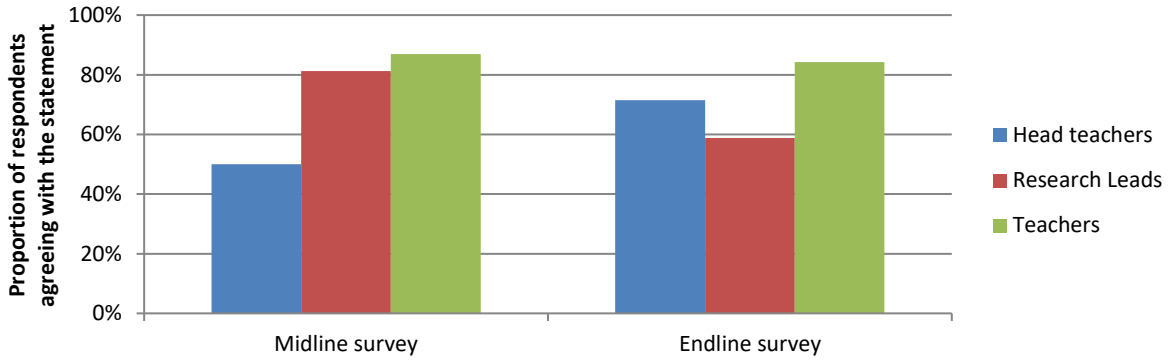
Another key element for the successful implementation of the Digital Feedback intervention is the use of diagnostic assessments. Almost all respondents in the midline and endline surveys, regardless of whether they were headteachers (100%: 7 at midline; 8 at endline), Research Leads (midline: 93%, 15 out of 16; endline: 88%, 15 out of 17) or teachers (midline: 100% (23); endline 95%,18 out of 19), reported (or self-reported) that teachers used diagnostic assessments (responded 'No' to 'I/teachers do not use diagnostic tests'). When asked in more detail in the midline and endline surveys about how often they were used, the majority of teacher and Research Lead respondents reported that teachers used diagnostic tests at least once a week or several times a week (midline: 75%, 12 out of 16; endline: 70%, 12 out of 17). However, reports on the ways in which the results from diagnostic assessments were used in daily teaching practice varied greatly, depending on participant role (headteacher, Research Lead, or teacher) and survey (midline or endline), as shown in Figure 12, Figure 13, and Figure 14.

Regarding teachers' use of diagnostic assessments, although there was variation in how participants responded—depending on role and time-point—overall, the majority of participants agreed or strongly agreed with the statement: 'Teachers use diagnostic assessments to focus teaching on areas that the class as a whole seem to be struggling with' (midline: headteacher, 50%, 4 out of 8; Research Lead, 81%, 13 out of 16; teacher, 87%, 20 out of 23; endline: headteacher, 71%, 5 out of 7; Research Lead, 60%, 10 out of 17; teacher, 84%, 16 out of 19—see Figure 12). Illustrating this point, the interviewed teacher stated:

[14] After completion of the trial, the delivery team informed the evaluation team that RLs were invited to attend one of the PLC meetings. However, this was not done systematically across the different PLC groups.

'The diagnostics were great because there were questions for every part of the curriculum […] you could have a question that assessed the key skills for the children, and that would then tell you if the children knew the skill or not. So if they didn't, you could then teach them. If they did, you could set them different questions' (INT).

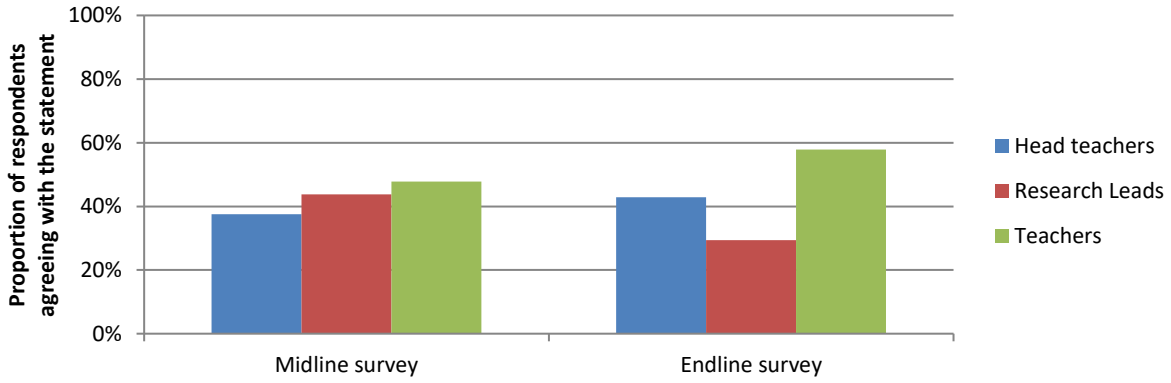
Figure 12: Survey responses for the statement 'Teachers use diagnostic assessments to focus teaching on areas that the class as a whole seem to be struggling with'



Midline: headteacher, n = 8, Research Lead, n = 16, teacher, n = 23; endline: headteacher, n = 7, Research Lead, n = 17, teacher, n = 19.

In comparison, agreement was generally lower for whether teachers used diagnostic assessments to track progress made by the class as a whole over the year (Figure 13). It is interesting to note that Research Leads and teachers reported similar levels of agreement at midline (teachers 48%, 11 out of 23; Research Leads 44%, 7 out of 16) only for views to diverge at endline, with Research Leads' level of agreement decreasing while teachers' level of agreement increased (teachers 58%, 11 out of 19; Research Leads 29%, 5 out of 17). In our observation of PLC meetings, teachers also reported that the diagnostic tests allowed teachers to see pupil progress through time at a fine-grained level (PLC3). As noted earlier, Research Leads appeared to have less direct contact with the intervention than teachers (that is, Research Leads reported lower training attendance and PLC meetings were only available to teachers) and therefore their views may not have been as directly informed by hands-on experience.

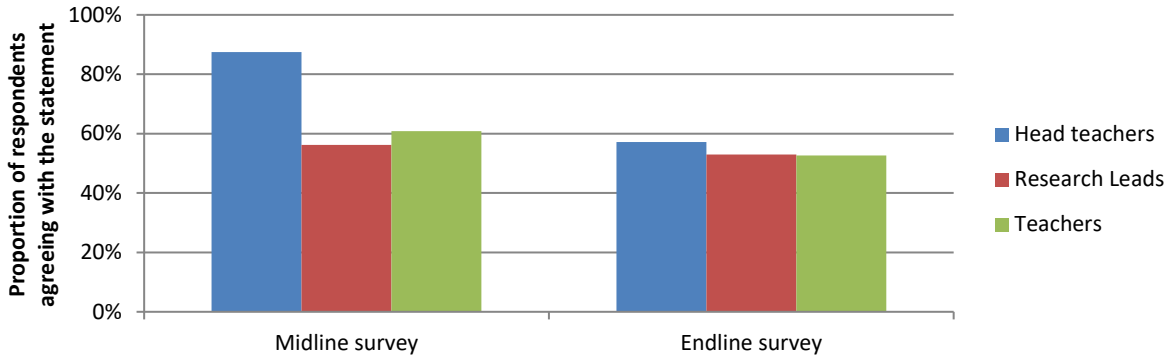
Figure 13: Survey responses for the statement 'Teachers use diagnostic assessments to track progress made by the class as a whole over the course of the year'



Midline: headteacher, n = 8, Research Lead, n = 16, teacher, n = 23; endline: headteacher, n = 7, Research Lead, n = 17, teacher, n = 19.

Regarding teachers' use of diagnostic assessments to target pupils with greater learning needs, the proportion of respondents agreeing or strongly agreeing with this statement decreased for all participant roles between the midline and endline surveys (midline: headteacher, 88%, 7 out of 8; Research Lead, 56%, 9 out of 16; teacher, 61%, 14 out of 23; endline: headteacher, 57%, 4 out of 7; Research Lead, 53%, 9 out of 17; teacher, 53%, 10 out of 19—see Figure 14). Taking the results reported in this sub-section as a whole, it may be that during the course of the intervention, participants used Digital Feedback less as a tailoring or targeting tool and more as a global classroom tool. This trend appears also to be reflected in how teachers used video feedback (as reported in the next section). It is important to highlight the caveat that as the surveys did not track respondents across time-points, the trends reported in this section should not be taken as longitudinal. Nevertheless, the findings articulate potential changes in general in the views participants may have held about using Digital Feedback and its linked components.

Figure 14: Survey responses for the statement 'Teachers use diagnostic assessments to target pupils with greater learning needs'



Midline: headteacher, n = 8, Research Lead, n = 16, teacher, n = 23; endline: headteacher, n = 7, Research Lead, n = 17, teacher, n = 19.

Implementation and fidelity—key takeaways

While bearing in mind survey data limitations, results from the IPE suggest that a noticeable proportion of teachers (endline survey) provided digital feedback less than once a month and that there were difficulties experienced in embedding Digital Feedback within classroom practice. Together, these suggest that there was poor compliance and a low dosage of digital feedback provided during the trial. This is of particular importance in light of the fact that the digital element of the intervention was one of its main innovations. However, it is also true that the decision as to *how much* digital feedback should be provided was largely left to the discretion of teachers—the suggestion was to use it two to three times per week depending on the needs of the class, but this was not mandated and of course teachers have the agency to follow or disregard suggestions. Furthermore, it is important to note that these survey responses only reveal that teachers were not providing *digital* feedback as often as it may be desired. This does not mean that teachers were not providing *any* feedback to their pupils directly—even if not using the app—and does not tell us whether they were providing more effective feedback as a result of the training received as part of the intervention.

Furthermore, the fact that teachers reported introducing adaptations to the intervention, that they had irregular coaching sessions with Research Leads (if any), and that there was irregular attendance to PLC meetings provides evidence of low fidelity to the intervention throughout the trial. Furthermore, the IPE data highlighted that many Research Leads did not attend the training that was available to them and many may have been overburdened by their SLT roles, which provides a potential reason for why coaching sessions were not held. These identified implementation caveats combined may serve as a plausible explanation for the null effects identified in the impact evaluation.

IPE: perceptions of effects on outcomes

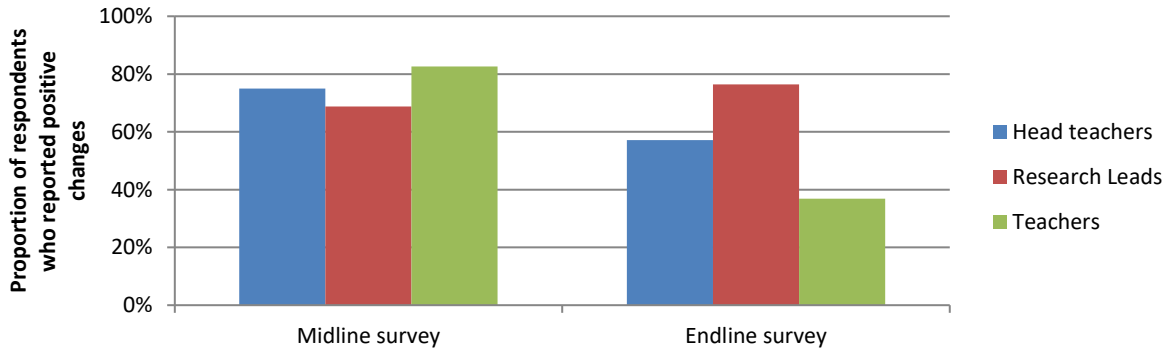
Box 7: Outcomes—key findings

- Survey respondents (teachers, Research Leads, and headteachers) expressed diverging views on (1) the perceived impact that Digital Feedback had on teacher practice and (2) whether the quality of feedback given to pupils increased as a result of the intervention. Research Leads remained more positive across time than teacher and headteacher survey respondents.
- There was a decrease between the midline and endline surveys in the proportion of teacher and headteacher respondents agreeing or strongly agreeing that there were positive changes in teacher feedback practices and improvements in the quality of feedback given to pupils as a result of the intervention.
- By the end of the trial more than half of teacher and Research Lead survey respondents agreed or strongly agreed that pupils engage more with digital feedback than with traditional written feedback. Anecdotal evidence (interview and open-ended survey responses), albeit small in sample size, suggest that Digital Feedback may have had some positive impact on pupils.
- The proportion of teacher and Research Lead respondents agreeing or strongly agreeing that pupils get distracted by digital feedback and do not pay attention to the content increased between the midline and endline surveys.

Perceived outcomes of using Digital Feedback in Primary Maths

The surveys revealed mixed feedback on the perceived impact of Digital Feedback on teachers. Teachers, Research Leads, and headteachers were asked whether there had been positive changes in the content of feedback provided to pupils since the beginning of the school year. While the proportion of Research Lead respondents agreeing that there had been positive changes increased between the midline (67%, 11 out of 16) and endline (76%, 13 out of 17) survey, for teachers (midline 83%, 19 out of 23; endline 37%, 7 out of 19) and headteachers (midline 75%, 6 out of 8; endline 57%, 4 out of 7) it was the opposite, with the proportion reporting positive changes decreasing (see Figure 15).

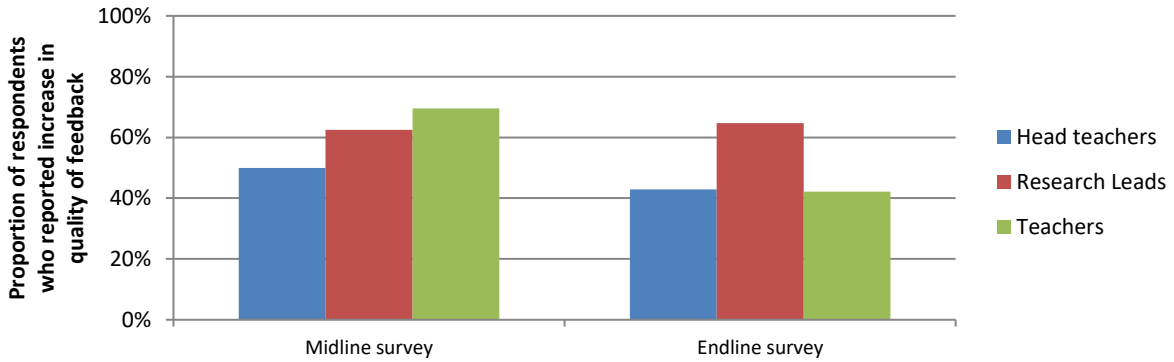
Figure 15: Positive changes in content of feedback given to pupils by intervention teachers, as reported by survey participants



Midline: headteacher, n = 8, Research Lead, n = 16, Teacher, n = 23; Endline: headteacher, n = 7, Research Lead, n = 17, teacher, n = 19.

A similar pattern emerged when respondents were asked whether the quality of feedback that teachers provided to pupils had increased as a consequence of using Digital Feedback. Again, while the proportion of Research Lead respondents reporting an increase in the quality of feedback increased slightly between the midline (63%, 10 out of 16) and endline (65%, 11 out of 17) survey, for teachers (midline 70%, 16 out of 23; endline 42%, 8 out of 19) and headteachers (midline 50%, 4 out of 8; endline 43%, 3 out of 7) the proportion reporting an increase in feedback quality decreased between the two surveys (see Figure 16).

Figure 16: Increase in quality of feedback given to pupils as a result of using Digital Feedback, as reported by survey participants



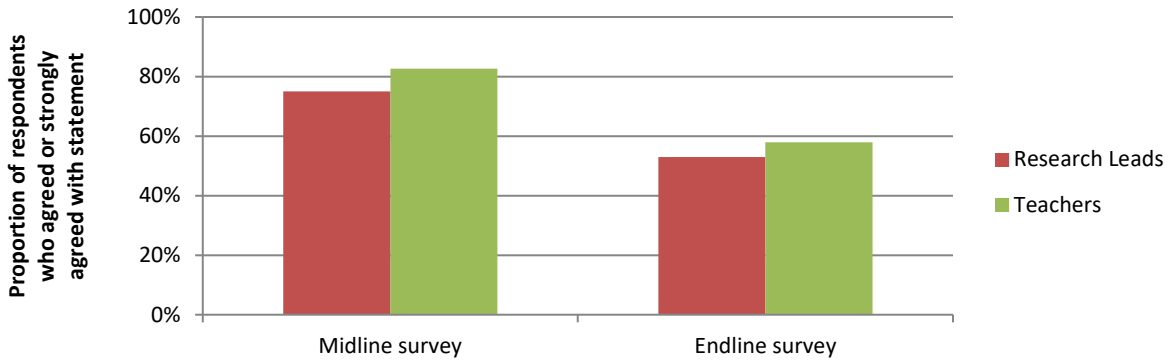
Midline: headteacher, n = 8, Research Lead, n = 16, Teacher, n = 23; Endline: headteacher, n = 7, Research Lead, n = 17, teacher, n = 19.

The surveys also revealed mixed feedback on the perceived impact of Digital Feedback on pupils. At the beginning of the intervention, Research Lead participants were very positive about the potential of Digital Feedback, with almost all respondents in the baseline survey (94.4%, 17 out of 18) stating that they expected the intervention to be able to contribute to pupil outcomes.

Teachers and Research Leads were asked the extent to which they agreed that that pupils engaged more with digital feedback than traditional written feedback. The proportion of respondents agreeing or strongly agreeing with the statement was at least 75% in the midline survey (teachers, 19 out of 23; Research Leads, 12 out of 16); in the endline survey, the proportion of teacher and Research Lead respondents agreeing or strongly agreeing with the statement was smaller—58% (11 out of 19) and 53% (9 out of 17) respectively (Figure 17). Despite the slight decrease in the

proportion of respondents agreeing with the statement, these findings show that by the end of the trial more than half of respondents agreed or strongly agreed that pupils engage more with digital feedback than with traditional written feedback.

Figure 17: Survey responses for the statement 'Pupils engage more with digital feedback than with traditional written feedback'



Midline: teacher, n = 23, Research Lead, n = 16; endline: teacher, n = 19, Research Lead, n = 17.

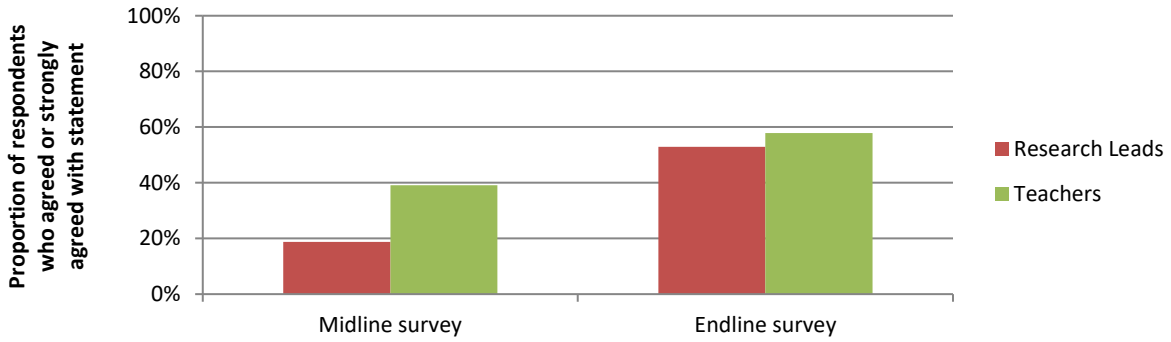
Nevertheless, there was anecdotal evidence that Digital Feedback may have had a positive impact on pupils. In surveys, some respondents reported, for instance, that '[Digital Feedback] has engaged pupils' (SUR teacher, midline), '[pupil] outcomes have improved' (SUR headteacher, endline) and that pupils were 'becoming independent and working on misconceptions themselves using the iPads' (SUR RL, midline). The teacher who was interviewed noted that pupils understood their mistakes better, as well as how to improve their mistakes:

> *'I think it helped them [pupils] be aware of mistakes they were making, that they would then be better at improving mistakes during the maths lesson. I think that was a valuable lesson' (INT).*

Perceived unintended consequences of using Digital Feedback in Primary Maths

One potential unintended consequence survey respondents were asked to report on was whether pupils were distracted by Digital Feedback and did not pay attention to the content of feedback. In the midline survey, a minority of Research Lead (18%, 3 out of 16) and teacher (39%, 9 out of 23) respondents agreed or strongly agreed that Digital Feedback was distracting for pupils; however, in the endline survey, the majority of respondents agreed or strongly agreed that pupils were distracted and did not pay attention to the content of Digital Feedback (Research Lead, 53%, 9 out of 17; teacher 58%, 11 out of 19—see Figure 18). Indeed, in the endline survey, one teacher respondent suggested that in order to be more effective, pupils needed the Digital Feedback application to be well embedded within class practice 'so the novelty doesn't distract others' (SUR teacher endline). Given survey limitations, these results may be influenced by selection bias, however, they could indicate that IT innovation may have a short-term 'pull' for pupils but they then focus on the IT itself and not the content.

Figure 18: Survey responses for the statement 'Pupils get distracted by digital feedback and do not pay attention to the content'



Midline: teacher, n = 23, Research Lead, n = 16; endline: teacher, n = 19, Research Lead, n = 17.

Formative findings

Box 8: IPE formative findings summary

- IPE data suggests that:
  - The training could be improved in terms of content: it should focus more on embedding Digital Feedback in the classroom—its effective use in practice. With regard to timing: only one training session prior to implementation was considered insufficient and that it covered too much material.
    - Going forward it would be worth considering restructuring the training so that it is spread over multiple sessions and takes place closer to the start of the intervention. It would also be worth considering if and how the PLC coordinators should proactively supplement the training in PLC meetings, especially at the beginning of the intervention.
  - The clarity and focus of the intervention can be improved—the intervention is made up of multiple components. Some participants felt that the digital element of Digital Feedback was neither a core component nor an important element of the intervention, and was the part of the intervention that was hardest to put into practice.
  - In the future it would be important to consider how the different parts of the Digital Feedback intervention fit together, and reassess the importance of the digital element. More consideration should be given to how teachers could implement the digital aspect of the intervention as part of their daily classroom practice if this element is deemed to be crucial.
  - Based on training observations, survey responses, PLC observations, review of RAG ratings, and coaching records, throughout the trial there were low levels of implementation fidelity and dosage of the different intervention components.
  - Reports on perceived outcome of the intervention on teacher feedback practice and whether feedback quality improved were mixed. Anecdotal evidence suggests that Digital Feedback may have had some positive impact on pupils' awareness of weaknesses and ability to overcome mistakes.

According to survey responses, training and PLC meeting observations, the review of RAG ratings, and the interview and focus group, **one of the aspects of the Digital Feedback intervention that could be improved was the training provided, both in terms of its structure and its timing**. In terms of its focus, as discussed in the 'Participant attendance at training and their perceptions' section, we observed in Research Lead training observations that more focus could have been placed on the coaching element of the intervention and the importance of securing time for it. Furthermore, previously presented data suggests that the preparation Research Leads received to coach and support teachers in the use of Digital Feedback seemed to be insufficient. We also identified that coaching was irregular and in the experience of some teachers did not take place at all. Findings suggest that a potential factor driving this was that selected Research Leads also fulfilled senior leadership roles which limited their time availability for the intervention. Moreover, in teacher and Research Lead training sessions we observed that when discussing the digital element of the intervention insufficient focus was placed on how to use the application effectively in the classroom. The teacher interviewee also noted that it would have been helpful to receive more information in this regard:

> *'I think our PLC group really understood what needed to happen from Christmas onwards. So the first three months were just trying to get our heads around how to do it […] I don't know if it [the training] was detailed enough, or didn't give real-world experiences, or maybe it was just a new thing that needed time to get into class […]. If you could see what the books look like or you could have seen the physical thing they actually did, or go and watch a lesson on how it could be delivered, that could have been really handy. Because you just had to work out how to fit it into your teaching [it would have been helpful to have] more modelling of what needed to be done' (INT).*

Relatedly, **in terms of the timing of training, several teachers felt that one training session prior to implementation had been insufficient** because of the amount of information that needed to be provided and processed in this initial session (PLC3, FG, PLC RAG ratings). In addition, **the fact that the training took place in the summer meant that some of the content may have been forgotten prior to the beginning of the intervention when the school year restarted** (FG). Some teachers suggested that more continuous training may have been beneficial (FG). Our analysis of PLC RAG ratings found that at the beginning of the year, teachers were uncertain about how to use Digital Feedback and may have required additional training. Furthermore, based on survey responses, PLC observations, and the focus group, it seems that in the experience of some teachers, their Research Leads were unable to fulfil their coaching role satisfactorily (that is, by meeting regularly) due to their busy schedules since they tended to

fulfil senior leadership roles, and in some cases due to not mastering the use of Digital Feedback themselves. Going forward it would be worth considering restructuring the training so that it is spread over multiple sessions and takes place closer to the start of the intervention. It would also be worth considering if and how the PLC coordinators should proactively supplement the training in PLC meetings, especially at the beginning of the intervention.

A second formative finding related to the Digital Feedback intervention itself. In the midline and endline surveys, **several teacher, Research Lead, and headteacher respondents reported that they felt the intervention lacked clarity and focus**. In particular, informants felt that the main focus of the intervention was, in fact, how to deliver effective feedback to pupils and how to use diagnostic assessments to aid the delivery of feedback. Some participants felt that the digital element of Digital Feedback was neither a core nor an important element of the intervention and was the part of the intervention that was hardest to put into practice (Box 9).

Box 9: Opinions on the core elements of the Digital Feedback intervention—open-ended survey answers

' "Do it nows" and diagnostics were very good. Use of technology was trickier to put into practice' (SUR teacher endline).

'It appears that the focus is more on the diagnostics and the feedback, rather than use of ICT to deliver the feedback' (SUR RL midline).

'Digital feedback is a very specific element of feedback whereas the project itself focused on many other variables' (SUR RL midline).

'Any benefits from this trial come from the increase[d] awareness and appreciation for effective feedback—NOT [emphasis original] the digital aspect. The digital aspect of this project is difficult to implement in school and to sustain' (SUR RL endline).

'Session on effective feedback was the key session but the use of technology does not impact—it's the quality of the feedback [that is key]' (SUR head teacher midline).

'Like with all feedback, written or digital, the effectiveness of the feedback relies on teacher knowledge of the subject and ability to identify and robustly intervene' (SUR head teacher endline).

In the future it would be important to consider how the different parts of the Digital Feedback intervention fit together and reassess the importance of the digital element. Ultimately the intervention was about CPD and support for teachers on effective feedback practices, with the digital element being a platform for introducing those elements. If the digital element actually gets in the way of the CPD element then it might be worth re-considering the inclusion of this element. **If it was felt that the digital aspect *is* an important part of the intervention, more consideration should be given to how teachers could implement it as part of their daily classroom practice**. This may include consideration of how to monitor and review the amount of feedback that is recommended as part of the intervention as well as practical elements such as recording of feedback and the ease of using the application. Results also suggest that the way teachers implement Digital Feedback should receive stronger focus in the initial training.

A third formative finding relates to implementation fidelity. As noted in the 'Implementation and fidelity—key takeaways' section, overall, throughout the trial process evaluation **we identified low levels of implementation fidelity and dosage of the different intervention components**. As already noted, meetings between Research Leads and teachers were irregular, similar to the attendance at PLC meetings. Still, despite irregular attendance, those joining PLC meetings reported these were useful. In terms of the use of Digital Feedback, IPE data suggests that the intervention was used more as a universal classroom tool rather than a targeting tool to support individual or struggling pupils. In terms of the amount of *digital* feedback provided by teachers, survey responses suggest that there was a low dosage of digital feedback provided by teachers during the trial (in the endline survey the majority of teacher respondents reported providing immediate and delayed feedback less than once a month). Other forms of feedback (for example, verbal or written) may have been given but not necessarily using the app.

Finally, another formative finding relates to participants' perceptions of outcomes resulting from the intervention. **Views on the outcomes of the intervention on teacher feedback practice and whether feedback quality improved were mixed**, although lower proportions of respondent agreed that there were improvements in the quality of their feedback in the endline survey compared to the midline survey. In terms of pupil outcomes, a majority of survey respondents agreed or strongly agreed that pupils engage more with digital feedback than with traditional written feedback.

Anecdotal evidence, albeit sparse, suggests that Digital Feedback may have had some positive impact on pupils (for example, pupils being better equipped to identify and improve on mistakes). This is in contrast to survey results where higher proportions of respondents agreed or strongly agreed than disagreed with the statement that pupils got distracted by technology.

Control group activity

Box 10: Control group activity—key findings

- Findings suggest that the control group may have experienced or implemented practices similar to intervention teachers.
- Control group survey respondents reported receiving coaching. A third reported having received coaching that focused on how to provide effective feedback.
- Nearly half of control group teacher survey respondents reported changing the way they provide feedback throughout the trial year (some reported providing more verbal feedback or more immediate feedback).
- Findings suggest that there was a high risk of contamination between control and intervention participants due to trial design and dual roles some participants were allocated (for example, being both Research Lead and control teacher).
- Observations of the teacher and Research Lead training suggest that training information on the risk of contamination could have been given greater prominence.
- Some intervention teachers reported that schools and fellow teachers started changing wider teaching practices as a result of Digital Feedback.

This section focuses on what 'business as usual' looked like in the control classes, as well as contamination in the trial. These are of particular interest as all schools included both control and treatment classes. Overall, the data collected suggests that the control group may have experienced or implemented practices similar to those in Digital Feedback (such as coaching or providing immediate feedback), whether due to changes in school practices or potential contamination.

Business-as-usual practice for control teachers

Amongst control teachers, two-thirds of respondents in the endline survey reported that they had not received coaching on how to provide effective feedback, while one-third of respondents reported that they had received such coaching. When asked to describe the coaching, control teacher respondents reported that coaching took place as part of staff meetings, team work with other teachers, peer-to-peer coaching, as well as in-class coaching. Data collected from teachers and Research Leads in the intervention group also indicated that in some schools, school-wide coaching programmes already existed (PLC2, SUR RL, midline; SUR RL, endline). Indeed, a minority of Research Lead respondents in the midline (18.8%) and endline surveys (23.5%) reported providing coaching other than to teachers in the treatment group—though this does not necessarily mean they coached teachers in the control group. Some Research Lead respondents pointed out that they were mindful of avoiding contamination as shown in Box 11.

Box 11: Comments on contamination by control participants—open-ended survey answers

'Obviously the Digital Feedback is not discussed with the control teacher' (SUR RL midline).

'All teachers are coached at a minimum of once per half term and the focus is on development of their teaching and is bespoke to their personal teaching journey. Apart from the digital [feedback intervention] teachers, no other teachers were coached on iPad technologies' (SUR RL endline).

Nevertheless, the data provides provisional evidence that coaching, a key element of Digital Feedback, was experienced by some teachers in the control group. In one PLC meeting, an intervention teacher reported that they had shared the readings provided by PLC coordinators with senior staff in their school in order to improve their coaching (PLC2). In addition, one teacher informant in the focus group explained that the school had implemented—alongside Digital Feedback—a different maths intervention with overlapping elements to Digital Feedback (FG).

In the endline survey, teachers in the control group were asked if they had changed the way they provided feedback to pupils since the beginning of the school year. A total of 41.7% of respondents reported that they had changed their feedback practices. These respondents reported that they provided more 'verbal' or 'oral feedback', as opposed to written feedback, as well as 'in-class', 'immediate feedback'—again pointing to potential similarities to Digital Feedback. When asked, 80% of these respondents agreed that the quality of feedback they provided had increased as a consequence of coaching they had received.

Finally, control teacher respondents reported using usual classroom assessment practices during maths lessons—for example, marking (100%), question rounds (83.3%), half-termly assessments (58.3%), weekly assessments (58.3%), as well as pop quizzes and Assessment for Learning strategies. Furthermore, the delivery team noted that all classes in selected year groups were given access to diagnostic assessments, meaning that control teachers may have also had access and potentially used this resource. As noted above, the use of diagnostic assessments was one of the most useful elements of the intervention (as noted by intervention teachers). This presents a potential limitation. However, the evaluation team is unable to assess the extent this may undermine the conducted analysis given that control teachers were not explicitly asked about the use of diagnostic assessments.

Potential contamination

As noted in the Methods section, during randomisation the research team had to ensure each school was allocated to at least one treatment class. In practice, this meant that each school was allocated at least one treatment and one control class. This increased the risk of contamination from the outset of the trial. Contamination was one of the topics addressed in the initial training for both teachers and Research Leads. However, our assessment from observing a teacher and Research Lead training session was that information on contamination had received lower priority than other topics (as detailed in the 'Participant attendance at training and their perceptions' section on page 39). In the teacher training, for instance, we observed that teachers were not given an explanation as to why contamination would be a problem for the trial, and the risks of control and treatment teachers planning lessons together were also not clearly presented. In the Research Lead training, the slides on contamination were presented only after some of the attendees had left.

Nevertheless, all of the Research Lead respondents in the midline survey who had attended the initial training agreed or strongly agreed that it was made clear to them during the training who should receive the intervention in their school (100%), what the risks were of contamination of the intervention (90.9%), and what their role was in preventing contamination of the intervention (90.9%) (RL midline Q12). In the midline and endline surveys, teacher respondents were asked if they planned lessons together with other teachers, and if so, if they knew what they should or should not do when planning lessons together with a teacher of a control class. The majority of teacher respondents reported that they either knew what they should or should not do, or that they did not plan lessons together with control teachers (midline: 90.9%; endline: 91.7%). Only one respondent in each of the surveys reported that they did not know what they should or should not do, or that they did not remember what the rules were.

However, contamination risk in the trial remained high, especially in the 14 schools where individual school staff served multiple roles (as headteacher, Research Lead, and/or control or treatment teacher) within the trial. Of particular concern to the evaluation team were the schools in which a staff member fulfilled a role in the treatment and the control arms simultaneously, for instance, acting as Research Lead and control teacher—which was the case in seven schools (21%) out of the 34 participating in the trial. Several of the Research-Lead control teachers (those with dual roles)[15] had either been trained as Research Leads or had attended PLC meetings. The evaluation team flagged the issue of multiple roles in November 2017 to James Siddle, the developer, who then contacted the relevant schools and, for instance, encouraged the school to elect a different Research Lead. Although the evaluation team was not able to definitively identify if these control teachers in the affected schools had used parts of the Digital Feedback intervention (for example, the diagnostic assessments or providing video feedback), it is possible that they may have integrated some elements of the intervention into their teaching practices.

[15] Based on records on the number of participants with dual roles from November 2017 there were:
five acting as headteacher and Research Lead;
seven as Research Lead and control teacher; and
one as Research Lead and treatment teacher.

During the evaluation team's observations of PLC meetings, some intervention teachers reported that their schools and fellow teachers had started changing wider teaching practices as a result of Digital Feedback (PLC1, PLC2). Some teachers felt that it was almost inevitable that schools would begin to adopt the good practice elements of the intervention (PLC1). One teacher expressed concerns about this creating more work as they had to do more to go beyond what was already being done by the control class (PLC2). Teachers were aware that they should prevent contamination as much as possible: for instance, one teacher explained that they tried to keep Digital Feedback 'a secret', and hide what they were doing whenever someone came into their classroom in case it was a teacher from the control group (PLC1). However, due to the potential contamination, they also expressed concerns that no differences on mathematics outcomes would be found between pupils in the intervention and control groups (PLC1).

Conclusion

Key conclusions	
1.	There is no evidence that Digital Feedback in Primary Maths had an impact on pupils' maths outcomes. This result has a moderate to low security rating.
2.	There is no evidence that the programme has an impact on the maths outcomes of pupils eligible for free school meals.
3.	The programme, as designed, was not easily implemented by teachers and schools. There was evidence that teachers and Research Leads sometimes did not attend training, have coaching sessions, or provide regular digital feedback in the classroom.
4.	Both the diagnostic assessment and training in effective feedback were reported as more useful by teachers than the use of the app to provide digital feedback. Evidence from surveys, PLC observations, and an interview indicated that the diagnostic assessments were eventually being used as a global classroom assessment tool.
5.	There were significant challenges with identifying which pupils were in the treatment group and which were in the control. This meant that 431 pupils could not be included in the main analysis. This presented a small threat to the security of the trial.

Interpretation

Teachers provide feedback on a daily basis as a means to remind, correct, or encourage their pupils. Approaches to making that process more efficient or effective, if successfully scaled, could have meaningful impacts on the daily lives of both teachers and pupils. Indeed, the EEF's own toolkit sets out that effective feedback is associated with up to eight months' progress (EEF, 2018c). This trial attempted to assess the efficacy of a digitally-integrated approach to providing feedback to pupils. The intervention was, in essence, an approach with elements of CPD that incorporated a 'digital' element through the use of tablets so that teachers could record feedback on maths tasks, but the core of the intervention was to train teachers on how to give better feedback. The intention being that through giving and recording *better* feedback, pupils could review explanations of an answer at their own pace, share with other pupils so that the same question would not need to be asked by the same pupil/group, and build up a repository of questions and answers that would provide a resource for revision. Whilst seemingly simple in its formulation, Digital Feedback in Primary Maths was a complex intervention that combined elements of CPD, reflective practice, coaching, 'communities of practice', 'Edtech', and a need for computing competence. The intervention required that schools had the IT resources to implement it, classroom teachers were appropriately trained, and that senior staff were available to coordinate, coach and take a lead on monitoring.

Digital Feedback as implemented in this trial did not lead to improvements in maths attainment for pupils in the trial compared to control classes. There was a consistent albeit non-significant difference in the level of disengagement, with treatment pupils reporting being less disengaged than controls. On its own, the disengagement result would not be sufficient to justify further implementation of this approach.

What the evaluation found was that although there were examples of good implementation, this was not consistent overall (for example, there were instances where adaptations to the intervention were reported) and overall implementation was poor. Further, initial enthusiasm about Digital Feedback appears to have waned over time, suggesting a need for 'booster' training or inputs from developers to maintain enthusiasm. It also seems that some teachers who attended training were not always fully informed about the approach being offered. Further, there were differences of opinion regarding the usefulness of training on feedback, compared to teachers being prepared to deliver *digital* feedback. Similarly, there was evidence suggesting the feedback app may have been difficult to use, although that was by no means a uniform view. Overall, our interpretation of the IPE results is that Digital Feedback is an approach that was not implemented as intended and which had some way to go before being evaluation-ready. In particular, the IPE data—sparse as that might be in places—suggests that there was an insufficient link between the various components of the approach and that the initial enthusiasm regarding the 'digital' element was not maintained over time. This is not to detract from the enthusiasm of the development team—who marshalled enormous effort to deliver the training—but more to argue that it would have been more suitable to commission this as a pilot project,

reigning in expectations about impact and focusing more on understanding the intervention with a view to future refinement and development.

From a research perspective, the trial faced many challenges as we have spelt out, but perhaps the biggest problem was 'designed in' from the outset and one that was insufficiently dealt with during the initial phases of the project. In short, the within-school design, coupled with Research Leads being based in the same school (and thus accessing both treatment and control teachers), meant that some control teachers reported that they too had received feedback training and that there was ample opportunity for contamination to routinely occur. Whilst we attempted to monitor this to an extent, the poor survey response rate, even to initial surveys, should have been read as a signal that schools were not sufficiently engaged in the research element of the project (and perhaps the intervention). As such, the overall 'no difference' result from the impact evaluation is perhaps unsurprising, particularly in the light of broader research illustrating the lack of impact of education trials (Lortie-Forgues and Inglis, 2019). That we also did not find any benefits for poorer pupils and were not able to assess effects for SEN pupils due to lack of data only speak to how wide the remaining gaps in knowledge are in this area.

In spite of these results there is a great deal that can be learnt from this trial and others like it. In particular, our view is that technological solutions to existing problems in teaching require significant investments of time understanding and appreciating 'user experience' and utility in education settings before they are combined with more traditional teaching or training (OECD, 2015; Bouygues, 2019). Using Digital Feedback as an example, we observed that even schools that were eligible for the study—meaning they had the IT infrastructure required—struggled to fully utilise the approach. But the shortfall was not, we believe, in enthusiasm; it was the gap arising from the application of a digital tool to an existing problem at a scale at which it had not been used before and without an implementation guide/training that facilitated that. Whilst it may be possible for someone with mastery of a particular app to demonstrate this during training, teachers in the study reported that those with responsibility for training on the app (Research Leads) sometimes did not know how to use it themselves very well. If we were to speculate, based on our study, it would be that the digital element of the intervention may have been presented and thus treated as secondary from the outset. For a digital approach like this to be used, it has to be simple and easy to integrate into classroom practice, and moreover teachers need to immediately be able to see the value of the approach for them and see where and how they will use it.

Finally, we also think there are elements of the intervention that could be considered for further development. First, it seems sensible to explore the possibility of using the core 'feedback' training that was developed as an intervention in its own right. The fact that some teachers reported they were happy with the training they received and found it useful, but that did not extend to their feeling confident in delivering Digital Feedback, suggests that the core training could be developed further. Second, the diagnostic assessment that was implemented as part of the digital element of the intervention seems to have been relatively well received and understood by teachers. What we do not know is if there are general principles underlying that approach that could be investigated with a mind to further development.

Limitations

Impact evaluation limitations

As we have set out in the report, there were several limitations to the impact evaluation: (1) difficulty linking pupil data to teachers/allocation, (2) contamination—as this was a within-school design, and (3) poor implementation and fidelity, as well as apparent adaptations.

At the early stages, the compromise in design from school, to year group, to class-level randomisation necessitated by recruitment problems should have been a hard stopping point for recruitment and the project overall. The better strategy would have been to delay the project start to build up recruitment over a longer time-frame. Pursuing the compromise design likely contributed to what was the larger problem for the study, namely, data from around 400 pupils in the largest schools in the study could not be included in the main analysis. This was because of problems with data collection during the study meaning that the treatment/control allocation could not be reliably linked to all pupils. As this was a loss at school level, the threat to the internal validity of the trial should be lessened as both treatment and control classes were lost, and this should otherwise be unrelated to the outcome. Relatedly, we also think the evaluation was hampered by the timing of the request for data on pupils and classes from schools. We know now, and would recommend, that any class-level randomisation within schools would require that all pupil data for a given year group was requested in

order that teacher-pupil linkage could be completed on the basis of the actual teacher teaching a given class. Furthermore, as noted in the 'Randomisation' section (page 14), some schools also informed us after randomisation that they were not happy with their allocation(s) so some of the withdrawals were likely for that reason. This may have resulted in some bias in attrition. Based on this experience, our recommendation is that when planning within-school randomisation, or in situations where (primary) schools are being asked to supply pupil data in advance of deciding class composition, requesting UPNs for *all pupils in a given year* and only retaining data for those eventually taught by study teachers would be a better approach. Doing so would mean that the schools are not under pressure to pre-empt teacher allocations, and the expected changes in teachers between school years presents less of a problem for evaluators.[16]

The within-school design significantly increased the risk of contamination. Despite survey participants reporting having clarity on issues around contamination (for example, the importance of not sharing or discussing intervention practices with control teachers), IPE data suggested that in some schools this happened. Apart from this, there were five schools where Research Leads were also allocated the role of control teacher.

The disengagement measure we used might have been too general, but in the absence of a maths-only questionnaire it had to suffice. With hindsight, at the pupil level it might have been better to ask questions about maths anxiety rather than school disengagement if the focus was really on having a secondary outcome. However, disengagement with school was proposed as a mechanism for the intervention so measuring that in general terms may have been appropriate. We otherwise believe the outcome data for ELM mathematics to be of good quality for the purposes at hand: the distribution suggests that the measure was able to differentiate pupil ability sufficiently well.

IPE limitations

The IPE evaluation suffered from poor response rates from schools to surveys, few schools providing materials for the case studies, and relatively few interviews being conducted. We could also have adapted IPE approaches to capture higher number of views of participants once it was clear that, for example, surveys had very low response rates—but equally we faced difficulties with access and engagement from schools more generally. This could have been improved with more contact with schools, but overall our view was that schools may not have been as aware of the research requirements during recruitment as they should have been. Finally, IPE activities did not closely explore the use of diagnostic assessments in control classes. We are therefore unable to assess how, if at all, the generalised access to diagnostic assessments (to all year group classes) may have impacted findings.

Generalisability of results

In general, interventions that are not implemented as intended are less likely to be effective. As such, if schools and/or developers are not able to successfully work together to get to the point where delivery is 'as intended' then null results are very likely. As such, the null result for ELM mathematics is generalizable to situations where implementation has not been as successful as planned or where interventions have not been developed sufficiently before testing. This is not intended flippantly: the risk is that without a better intervention development pipeline the 'nothing works' in education story may continue for a long time.

Furthermore, the study's external validity is limited by the intervention's sample size. As noted in the 'Sample size' section (page 13), recruitment problems meant that fewer schools took part in the trial than originally planned. In addition, generalisability of the study is limited by the nature of schools that were recruited. The KYRA Teaching School Alliance recruited some schools that were members of its network within the areas it operated but not beyond. Many of the schools may not have been network members but had links with the KYRA Alliance lead.

[16] It is still an issue if a previously randomised teacher leaves and is replaced prior to the intervention starting. On the one hand, this could be viewed as 'ignorable' but with a within-school design this might lead to imbalances if there are few teachers being randomised within each school. On the other hand, it might be that the new teacher should be re-randomised into treatment or control, briefing schools on that possibility at the recruitment stage.

Future research and publications

- Based on the findings above, developing and testing a feedback intervention based on the core feedback training offered seems reasonable. A question is whether that approach would differ substantially enough from existing feedback approaches, and that question would need further consideration.
- The diagnostic assessment tool seems to have been used more widely than anticipated. As such, perhaps that could be developed and tested as a stand-alone approach to help teachers quickly assess pupil difficulties in maths.
- Focusing on implementation requirements for anything 'digital' and fleshing out what ‘success’ looks like seems to be important for EEF trials and other evaluators. For example, ensuring that an approach is integrated fully into existing IT systems and has been 'classroom tested' to a sufficient degree to allow for guidance to be written that would allow implementation beyond the core delivery team.

References

Australian Institute for Teaching and School Leadership (2017) 'Reframing feedback to improve teaching and learning'. https://www.aitsl.edu.au/docs/default-source/research-evidence/spotlight/spotlight-feedback.pdf?sfvrsn=cb2eec3c_12

Bouygues H. L. (2019) 'Does Educational Technology Help Students Learn?', Forbes. https://www.forbes.com/sites/helenleebouygues/2019/06/14/does-educational-technology-help-students-learn/#3447f3174539

Crawford, C. (2018) 'What Metacognition Looks Like In The Primary Classroom: A 7-Step Practical Approach To Maths Teaching', Third Space Learning. https://thirdspacelearning.com/blog/7-steps-eef-metacognition-primary-classroom-maths/

Dawson, A., Yeomans, E. and Brown, E. R. (2018) 'Methodological challenges in education RCTs: reflections from England's Education Endowment Foundation', *Educational Research*, 60 (3), pp. 292–310.

Dong, N. and Maynard, R. (2013) 'PowerUp!: A tool for calculating minimum detectable effect sizes and minimum required sample sizes for experimental and quasi-experimental design studies', *Journal of Research on Educational Effectiveness*, 6 (1), pp. 24–67.

Education Endowment Foundation (2016) 'A marked improvement? A review of the evidence of written marking', EEF and University of Oxford. https://educationendowmentfoundation.org.uk/public/files/Publications/EEF_Marking_Review_April_2016.pdf

Education Endowment Foundation (2018a) 'Feedback'. https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit/feedback/

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

Education Endowment Foundation (2018c) 'Feedback and monitoring pupil progress: Using assessment to understand pupil's strengths, weaknesses and progress, and how to address these through feedback'. https://educationendowmentfoundation.org.uk/school-themes/feedback-monitoring-pupil-progress/

Education Endowment Foundation (2019) 'ReflectED (re-grant): Project summary'. https://educationendowmentfoundation.org.uk/pdf/generate/?u=https://educationendowmentfoundation.org.uk/pdf/project/?id=1712&t=EEF%20Projects&e=1712&s=

Glennerster, R. and Takavarasha, K. (2013) *Running randomized evaluations: A practical guide*. Princeton University Press.

Hattie, J. and Timperley, H. (2007) 'The power of feedback', *Review of educational research*, 77 (1), pp. 81–112. http://www.ncbi.nlm.nih.gov/pubmed/20078751.

Hattie, J. (2008) *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.

Hedges, L. V. and Hedberg, E. C. (2007) 'Intraclass correlation values for planning group-randomized trials in education', *Educational Evaluation and Policy Analysis*, 29 (1), pp. 60–87.

Higgins, J. P., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., Savović, J., Schulz, K. F., Weeks, L. and Sterne, J. A. (2011) 'The Cochrane Collaboration's tool for assessing risk of bias in randomised trials', *BMJ*, 343, p. d5928.

Hout, M. (no date), personal communication.

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R. and Kerr, K. (2016) 'Implementation and process evaluation (IPE) for interventions in education settings: An introductory handbook', London: EEF.

Hutchison, D. and Styles, B. (2010) *A guide to running randomised controlled trials for educational researchers*, Slough: NFER.

Ketterlin-Geller, L. R. and Yovanoff, P. (2009) 'Diagnostic assessments in mathematics to support instructional decision making', *Practical Assessment, Research and Evaluation*, 14 (16), pp. 1–11. https://pdfs.semanticscholar.org/d8d8/6d37e7bfdab328e35c75519b5ecbcb85318b.pdf

Lortie-Forgues, H. and Inglis, M. (2019) 'Rigorous Large-Scale Educational RCTs Are Often Uninformative: Should We Be Concerned?', *Educational Researcher*, p. 0013189X19832850. http://eprints.whiterose.ac.uk/141754/

Oakley, A., Strange, V., Bonell, C., Allen, E. and Stephenson, J. (2006) 'Process evaluation in randomised controlled trials of complex interventions', *BMJ*, 332 (7538), pp. 413–416.

OECD (2005) *Teachers Matter: Attracting, Developing and Retaining Effective Teachers* (Summary in English), Paris: OECD Publishing. https://doi.org/10.1787/9789264018044-en

OECD (2015) *Students, computers and learning: Making the connection*, Paris: OECD Publishing. http://www.oecd.org/education/students-computers-and-learning-9789264239555-en.htm

Sibieta, L. (2018) 'The teacher labour market in England: shortages, subject expertise and incentives'. https://epi.org.uk/publications-and-research/the-teacher-labour-market-in-england/

Siddle, J. (2015) 'Verbal and visual-digital feedback on creative writing improves progress rates compared to written feedback in rural primary schools – a preliminary study', KYRA Teaching School Alliance.

Sim, M., van Belle, J. and Sutherland, A. (2018) 'Trial Evaluation Protocol: Digital feedback in Primary, Version 2'. https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/KYRA_Protocol_2018.08.20_Protocol_Update.pdf

Skinner, E., Furrer, C., Marchand, G. and Kindermann, T. (2008) 'Engagement and disaffection in the classroom: Part of a larger motivational dynamic?', *Journal of Educational Psychology*, 100 (4), p. 765.

Skinner, E. A., Kindermann, T. A. and Furrer, C. J. (2009) 'A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom', *Educational and Psychological Measurement*, 69 (3), pp. 493–525.

Snijders, T. A. B. and Bosker, R. J. (2012) *Discrete dependent variables. Multilevel analysis: an introduction to basic and advanced multilevel modeling*, pp. 304–307.

StataCorp. (2017) *Stata Statistical Software: Release 15*, College Station, TX: StataCorp LLC.

Sutherland, A. and van Belle, J. (2018) 'Statistical Analysis Plan: Digital feedback in Primary Maths, Version 2'. https://educationendowmentfoundation.org.uk/public/files/Projects/KYRA_SAP_2018.08.20_UPDATE.pdf

Teach Thought Staff (2017) '50 Questions To Help Students Think About What They Think', Teach Thought. https://www.teachthought.com/critical-thinking/metacognition-50-questions-help-students-think-think/

Tobey, C. R. (2017) 'Identifying Student Misconceptions with Formative Assessment Math Probes', McGraw Hill Education. https://s3.amazonaws.com/ecommerce-prod.mheducation.com/unitas/school/explore/research/reveal-math-formative-assessment.pdf

Worth, J. (2018) 'Latest teacher retention statistics paint a bleak picture for teacher supply in England', National Foundation for Educational Research. https://www.nfer.ac.uk/news-events/nfer-blogs/latest-teacher-retention-statistics-paint-a-bleak-picture-for-teacher-supply-in-england/

Appendix A: EEF cost rating

Cost ratings are based on the approximate cost per pupil per year of implementing the intervention over three years. More information about the EEF's approach to cost evaluation can be found **here**. Cost ratings are awarded as follows:

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

Appendix B: Security classification of trial findings

OUTCOME: *Maths Attainment (ACER's Essential Learning Metric)*

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

Threats to validity	Threat to internal validity?	Comments
Threat 1: Confounding	Moderate	The evaluators were unable to confidently link 17% of pupils to the treatment or control group – although because of the within school design, this does mean this issue should be roughly balanced across the treatment and control group. This does not allow them to compare balance at baseline of randomised treatment and control pupils. See missing data box.
Threat 2: Concurrent Interventions	Low	Not reported as an issue.
Threat 3: Experimental effects and contamination	High	As a result of the within school randomised design, contamination risk in the trial is high. In 14 of the 34 schools, individual school staff served multiple roles (as head teacher, Research Lead and/or control or treatment teacher) within the trial. Of particular concern to the evaluation team were the schools in which a staff member fulfilled a role in the treatment and the control arms simultaneously, for instance, acting as Research Lead and control teacher (which was the case in 7 schools (21%) out of the 34 participating in the trial).
Threat 4: Implementation fidelity	Moderate	The process evaluation provided some indication that the intervention was not delivered as intended. Although this data is not available for all participants, only slightly more than half of Research Lead respondents reported attending the half-day training (56.3%; 9 out of 16). Very few teachers reported giving digital feedback 'several times per week' (midline 4%, 1 out of 23; endline none out of 19) – although the exact amount was supposed to be left up to the teachers. Intervention survey responses showed a general decrease in the amount of digital feedback (immediate and delayed) given by teachers as the trial progressed. Reports on the use of diagnostic assessments may suggest that during the course of the intervention participants used Digital Feedback less as a tailoring or targeting tool as intended, and more as a global classroom tool.

Threat 5: Missing Data	Moderate	The evaluators were unable to confidently link 17% of pupils to the treatment or control group – although because of the within school design, this does mean this issue should be roughly balanced across the treatment and control group. This created a missing data issue on treatment status. They undertook sensitivity analysis on by imputing all pupils missing treatment allocation first to control, then doing the same but imputing to treatment. The results were the same as the complete case analysis.
Threat 6: Measurement of Outcomes	Low	No issues for primary outcome measure and sufficiently justified.
Threat 7: Selective reporting	Low	Not an issue in this trial.

- **Initial padlock score:** 4 padlocks. A within school randomised design with an MDES at randomisation of 0.187 SD. However; while the evaluators followed-up with 100% of randomised pupils for the primary outcome post-test data collection, because of issues with linking of pupils to treatment / control allocation (the evaluators were not confident of the treatment allocation of 431 pupils), they were only able to analyse the data of 83% of the pupils for the primary outcome.
- **Reason for adjustment for threats to validity**: - 2 padlocks. The report details various contamination risks include teachers in the control sample acting as coaches for intervention teachers (high risk), partially a result of the within school design. The issue with being unable to link the treatment status of pupils to outcome data effectively created an issue of missing data, although with their imputation approach to missing data analysis the results are the same as the complete case analysis (moderate risk). This may have introduced confounding to the estimates if there are some characteristics of the schools / pupils with missing allocation data associated with the outcome, and also did not allow the evaluators to assess imbalance of randomised pupils at baseline (moderate risk). In addition, the process evaluation provided evidence that the intervention was not delivered as intended (moderate risk).
- **Final padlock score:** initial score adjusted for threats to validity = **2 padlocks**

Appendix C: School Memorandum of Understanding

Memorandum of Understanding

Digital Feedback in Primary Maths

Introduction

This information is for schools wishing to take part in the 'Digital Feedback in Primary Maths' evaluation in the 17/18 academic year. **The aim of this trial is to increase the specificity and relevance of teachers' feedback using a digital app and tablet, making it easier for pupils to respond to it.**

The intervention will be delivered by staff from the Kyra School Alliance and Bishop Grosseteste University ("the delivery team"). The evaluation is being funded by the Education Endowment Foundation[1] and the evaluation is being undertaken by RAND Europe[2] ("the evaluation team").

The intervention is focused on Year 4 and Year 5.

Schools that agree to participate in the evaluation are being asked to sign this Memorandum of Understanding to ensure that the roles and responsibilities of the intervention and evaluation teams are clear to all partners and that schools understand what involvement in the evaluation means for them. **In signing the MOU, schools are confirming they have read and understood the requirements set out in the "Digital Feedback in Primary Maths evaluation information sheet" document that accompanies this MOU.**

The evaluation is a 'randomised controlled trial' (RCT)

In July 2017, year groups 4 and 5 within eligible schools that have completed the pre-requisites will be **randomly assigned to either 'control' or 'intervention' (akin to a 'coin flip')**. This means that each year group within a school has a 50:50 chance of being in the intervention condition, but it is the *comparison* between pupils in treatment and control schools that makes the evaluation possible and worthwhile.

Schools assigned to **intervention** will be provided with Digital Feedback in Primary Maths training sessions and take part in Digital Feedback in Primary Maths.

Schools assigned to the **control** condition will conduct the school year as 'business as usual'.

Eligibility criteria

To be eligible for inclusion in the evaluation, schools must meet all of the following criteria:

Eligibility criteria	Check (√) if condition met
Be a primary or junior school in England	
Sufficient number of tablets for each Year 4 and Year 5 classroom (roughly 1 tablet per 4 or 5 pupils)	
No prior engagement in Digital Feedback in Primary Maths	

If you are not sure if your school fulfils these criteria then please contact a member of the evaluation or delivery team to discuss (contact details at the end of this form). If your school is eligible and wishes to participate, **the first step is to get this memorandum of understanding signed by the Headteacher and at least one other member of the Senior Leadership Team (SLT)** and return it in person, via post, or scanned and emailed to the delivery team.

In order for a school to be officially part of the evaluation and considered for randomisation there are further requirements that must also be fulfilled (detailed below). (Note that pre-requisites relate *solely* to inclusion in the evaluation, not 'getting Digital Feedback in Primary Maths' since the intervention is allocated via 'coin flip' as noted above.)

[1] https://educationendowmentfoundation.org.uk/
[2] http://www.rand.org/randeurope.html; www.cam.ac.uk

1

If your school is eligible…read on…

Roles and responsibilities

There are three 'actors' involved in making this evaluation successful: the schools, the delivery team and the evaluation team. In order for the evaluation to work, each actor must understand their responsibilities and undertake their tasks as required. Schools are, of course, crucial to all of this so we (the delivery and evaluation teams) need to be as clear and up-front as possible about what schools need to do in the lead-up to, and when participating in, the evaluation.

The school's responsibilities are set out below – the roles and responsibilities of the delivery and evaluation teams are summarised here and detailed below. Following the description of the schools' role, we set out the pre-requisites for inclusion in the evaluation – tasks that schools need to complete before they can be part of the project.

School roles and responsibilities to enable the Digital Feedback in Primary Maths intervention to take place

It is important to keep in mind that the intervention consists increasing the specificity and relevance of teachers' feedback using a digital app and tablet. As such, participating schools will be expected to support the Digital Feedback in Primary Maths intervention in the following ways:

Selection of a Research Lead in each school. Each school will need to identify and select a Research Lead who will support the teachers in the school with the implementation of Digital Feedback in Primary Maths.

Enabling access to Digital Feedback in Primary Maths training. Training sessions will be offered to train teachers in the principles of effective feedback and in the use of the app. Sessions are offered for the teachers and Research Leads. Successful implementation requires attendance at these training sessions to ensure that the school has sufficient capacity to implement the approach.

Enabling access to Digital Feedback in Primary Maths Professional Learning Communities. All teacher assigned to the 'intervention' group will be asked to attend Professional Learning Communities meetings organised every four weeks.

School roles and responsibilities to enable the evaluation activities to take place

The evaluation consists of data collection from schools to investigate the effectiveness of Digital Feedback in Primary Maths. That data will consist of interviews, observations, surveys with teachers, Research Leads and head teachers, and information about pupils that allows us to: (a) request data from the National Pupil Database. Participating schools will be expected to support the Digital Feedback in Primary Maths evaluation in the following ways:

Completing the evaluation pre-requisites that are summarised below and detailed in the 'Digital Feedback in Primary Maths information sheet'.

Supporting the distribution of online surveys so that the views and experiences of teachers, Research Leads, and the head teacher in the school can be gathered.

Facilitating access to schools for evaluation team researchers so that they can undertake observations and/or interviews.

Facilitating access to schools so that outcome testing can take place at the end of the intervention, even if the school withdraws from the intervention during the school year.

Pre-requisites to be considered for randomisation

Once the MOU is signed, there are a few more steps to take before the school can be included in the randomisation (which is the point at which the school is officially 'in' the evaluation). These points are set out in the table below for information and we discuss them in more detail in the 'Digital Feedback in Primary Maths information sheet'. We have provided a collated checklist of both the eligibility criteria and pre-requisites as part of this information pack.

Pre-requisites to be considered for randomisation	Check (√) if condition met
Memorandum of understanding signed by Head + at least 1 other SLT member	
Parental opt-out forms distributed (and collected where applicable)	

Communicate to staff about the evaluation	
Pupil and teacher information shared with evaluation team	
"Baseline" teacher online survey distributed	

Data protection & anonymity

The evaluation team will collect and store all evaluation data in accordance with the Data Protection Act (1998). Evaluation data will be stored on secure servers. Data transferred between the delivery and evaluation teams containing any identifying information will be passphrase encrypted. Data transferred between schools and the evaluation team will be via encrypted files or secure file transfer.

Pupil data for supplying to the NPD will be transferred securely via the Department for Education secure transfer website (https://transfer.ecloud.education.gov.uk/submit).

No-one who participates in the evaluation will be named in any publications arising from the evaluation and results will be presented on aggregate. Schools will have the option of allowing their participation in the evaluation to be known.

Signature – Agreement for school to participate

In order to participate in this evaluation, schools must ensure that an MoU is signed by the Headteacher **and at least one other member** of the Senior Leadership Team (SLT).

By signing this memorandum of understanding we:

(i) **confirm that we have read and understood the eligibility requirements for this evaluation and that we fulfil all applicable eligibility requirements;**

(ii) **confirm that we have read and understood the pre-requisites for participating in this evaluation and that we commit to completing these in order to be included in the evaluation;**

(iii) **confirm that we have read and understood the document that sets out 'roles and responsibilities' for the evaluation team, the delivery team and schools;**

(iv) **agree to engage with the evaluation and delivery teams to facilitate the evaluation tasks;**

(v) **confirm that in the event that the school later wishes to withdraw from the evaluation activities the evaluation team will still be able to complete outcome tests with Year 4 and Year 5 pupils;**

_____

PRINT school name

Headteacher name (printed):	**SLT member name (printed):**
Headteacher signature:	**SLT member signature:**
Headteacher email address:	**SLT email address:**

Digital Feedback in Primary Maths eligibility and pre-requisite checklist for schools

For schools participating in the 'Digital Feedback in Primary Maths evaluation

Conditions for participation	**Tick (√) if condition met**
Eligibility criteria	
Be a primary or junior school in England	
Sufficient number of tablets for each Year 4 and Year 5 classroom (roughly 1 tablet per 4 or 5 pupils)	
No prior engagement in Digital Feedback in Primary Maths	
Pre-requisites to be considered for randomisation	**Tick (√) if condition met**
Memorandum of understanding signed by Head + at least 1 other SLT member	
Parental opt-out forms distributed (and collected where applicable)	
Communicate to staff about the evaluation	
Pupil and teacher information shared with evaluation team	
"Baseline" teacher online survey distributed	

Space for notes about the evaluation/questions/reminders:

4

Information sheet: Digital Feedback in Primary Maths

1. Introduction

The Kyra School Alliance (KSA) will take part in the Digital Feedback in Primary Maths trial and evaluation in the 2017/18 academic year. The aim of this trial is to increase the specificity and relevance of teachers' feedback using a digital app and tablet, making it easier for pupils to respond to it. The research is funded by the Education Endowment Foundation[3] and the independent evaluation will be conducted by RAND Europe[4]. Participating schools could benefit from increased pupil attainment in mathematics and improved diagnosis and feedback skills amongst teachers.

Teachers of Primary Years 4 and 5 can choose to take part. Before the commencement of the school year, each year group will be randomly assigned to a condition: Control, or Digital Feedback. Teachers in the Digital Feedback group will be provided with resources to quickly assess their pupils' understanding of a topic (for example, through multiple choice questions) in order to identify which pupils may be in need of further support. They will then begin providing feedback on mathematics tasks to these pupils by using tablets to record video summaries of their feedback, rather than writing down a comment. This will take two forms: immediate feedback, recorded and delivered on-the-spot in class to all pupils (depending on need); and delayed feedback, targeted specifically at pupils with a weak understanding of a particular topic, and which will replace written feedback. Pupils will have class time to review the feedback and develop as a result. Teachers in the Digital Feedback group will receive various forms of training and coaching throughout the year to support them in delivering effective digital feedback. Teachers randomised into the control group will not receive any coaching or training in digital feedback nor have access to the app, and will instead teach as "business as usual".

A mathematics test at the end of the academic year will measure pupil development as a result of the intervention. Surveys will also be conducted amongst teachers and designated school research leaders (RLs) at various points during the academic year to understand their experience of the digital feedback programme. Pupils will also be asked to complete an engagement test at the end of the year. The evaluation team may also conduct interviews with a small number of teachers, and review a sample of digital feedback videos.

2. Eligibility criteria

To be eligible for inclusion in the evaluation, schools must meet all of the following criteria:

Eligibility criteria	Check (√) if condition met
Be a primary or junior school in England	
Sufficient number of tablets for each Year 4 and Year 5 classroom (roughly 1 tablet per 4 or 5 pupils)	
No prior engagement in Digital Feedback in Primary Maths	

If you are not sure if your school fulfils these criteria then please contact a member of the evaluation or delivery team to discuss. If your school is eligible and wishes to participate, **the first step is to get this memorandum of understanding signed by the Headteacher and at least one other member of the Senior Leadership Team (SLT)** and return it in person, via post, or scanned and emailed to the delivery team.

In order for a school to be officially part of the evaluation and considered for randomisation there are further requirements that must also be fulfilled (detailed below). (Note that pre-requisites relate *solely* to inclusion in the evaluation, not 'getting Digital Feedback in Primary Maths' since the intervention is allocated via 'coin flip' as noted above.)

[3] https://educationendowmentfoundation.org.uk/
[4] http://www.rand.org/randeurope.html

3. Pre-requisites to be considered for randomisation

Once the MOU is signed, there are a few more steps to take before the school can be included in the randomisation (which is the point at which the school is officially 'in' the evaluation). These points are set out in the table below for information and we discuss them in more detail in the 'Digital Feedback in Primary Maths information sheet'. We have provided a collated checklist of both the eligibility criteria and pre-requisites as part of this information pack.

Pre-requisites to be considered for randomisation	**Check (√) if condition met**
Memorandum of understanding signed by Head + at least 1 other SLT member	
Parental opt-out forms distributed (and collected where applicable)	
Communicate to staff about the evaluation	
Pupil and teacher information shared with evaluation team	
"Baseline" teacher online survey distributed	

4. Data collection

We will collect data throughout the year (Table 1). In the following sections of this MoU we will outline in detail which parties will be involved with the different streams of data collection, and what their responsibilities will be.

TABLE 1: Data to be collected over the 2017-18 school year: Digital Feedback group

	At baseline	Mid year	At year end
Teacher survey (Digital feedback)	✓	✓	✓
Head teacher survey	✓	✓	✓
Research leader survey	✓	✓	✓
Pupil engagement test			✓
UPNs of classes (to be) taught in 17/18	✓		
School administrative data	✓		
External mathematics test			✓
Sample of Digital Feedback videos for review		✓	
Sample of diagnostic tests for review		✓	
Sample of coaching records for review		✓	
Interviews with teachers and research leaders			✓

TABLE 2: Data to be collected over the 2017-18 school year: Control group

	At baseline	Mid year	At year end
Teacher survey	✓		✓
UPNs of classes (to be) taught in 17/18	✓		
School administrative data	✓		
External mathematics test			✓

5. The role of the participating school

Schools that agree to participate in the Trial will be asked to sign this Memorandum of Understanding to ensure that the role and responsibilities of the intervention and evaluation are clear to all partners.

School role and responsibilities for the intervention

Participating schools will be expected to support the digital feedback intervention as well as coaching in the following ways:

- **Informing teachers** – communicate to all eligible teachers the opportunity to participate and the possible impact on their workload. Interested teachers will sign consent forms agreeing to participate.
- **Informing parents** – Schools will distribute informational letters (opt-out) about the evaluation to parents of children in participating classrooms. Pupils will be automatically enrolled in the evaluation unless their parent returns the information letter with an indication that they would like their child or children to *opt out* of the evaluation. Prior to randomisation, the school will provide a confirmation that the informational letter has been sent out and a list of pupils who have opted out, if any.
- **Access to tablets** – make available a sufficient number of tablets for use in the intervention classes (1 tablet for every 4 or 5 pupils as a minimum).
- **Access to training** – training about digital feedback will be provided to teachers and research leaders (RLs) in the summer before the commencement of the school year. Teachers from participating schools will attend these sessions to ensure that they have sufficient capacity to participate in the evaluation.
- **Access to internal and external coaching and support** – teachers participating in the intervention will be allowed to take part in monthly "professional learning communities" led by KSA and Bishop Grosseteste University, and access ongoing support and coaching from the designated school research leader in the form of a weekly 15-minute classroom observation and 20-30 minute meeting. Teachers will also develop monthly personalised action plans and targets to support their use of digital feedback. Coaching during the school year by the research leader will follow general guidelines, but each school and research leader can implement coaching and support as they wish. RLs themselves will meet once every term. For schools with a small number of staff, RLs may be drawn from another nearby school.

School role and responsibilities for the evaluation

In order to support the evaluation, schools will be asked to provide access to their teachers and pupils to KSA and RAND Europe for:

- **Teacher survey** – online surveys will be distributed among teachers in the Digital Feedback group at three time points: at baseline, in the middle of the academic year, and at the year end. The first survey will collect information on teachers' background and views on digital feedback. The second and third survey among teachers will focus on their experience of Digital Feedback and the use of the resources to assess pupil's understanding of a topic. Each survey should take between 10-15 minutes to complete. Teachers in the Control group will complete surveys at baseline and endline only.
- **Research leader surveys** – three RL surveys will be conducted during the school year to capture the views of the RLs on provision of digital feedback by teachers to pupils. RLs will be asked to give their view on how well and how often digital feedback is provided by teachers. Each survey should take between 10-15 minutes to complete.
- **Head teacher survey** – three surveys will be conducted during the school year to capture the views of head teachers on their perceptions of the implementation of Digital Feedback at their school, and the costs of implementation. Head teachers of schools in which all years are in the Control group will not fill out the survey.
- **Pupil engagement test** – an engagement test will be administered to pupils at the end of the year. This will ask pupils about their experience with the video feedback. Each engagement test should take between 10-15 minutes to complete.
- **Mathematics test** – a mathematics test will be administered at the end of the academic year to pupils in both groups (i.e. Control and Digital Feedback) by the Australian Council for Educational Research (ACER). The tests will be paper based and will take 45-50 minutes to complete. ACER will independently deliver,

invigilate and mark the tests. ***Schools will be asked to support the implementation of the external mathematics test even if teachers withdraw from the intervention.***

- **Collection of data** – RLs will keep records of the frequency and outcome of diagnostic tests administered to pupils; records of coaching sessions; and records of how well and how often the app is used by teachers to deliver feedback. This data will be made available to KSA.
- **Review of digital feedback** – The video feedback of a random sample of selected teachers will be reviewed by RAND Europe and tracked over time using a dedicated coding/observation protocol to measure improvement.
- **Review of diagnostic tests** – RAND Europe may analyse a selection of the diagnostics tests used in the classroom to asses pupils' understanding of a particular topics.
- **Review of coaching records** – RAND Europe may review a selection of RL coaching records to assess teacher engagement with the app.
- **Interviews** – a selection of teachers and research leaders from intervention schools will be invited for semi-structured in-depth telephone interviews to discuss their experience of the Trial. Participation will be voluntary.

In addition, schools will be asked to provide administrative and background data to support the evaluation. To conduct the evaluation, schools are asked to consent to sharing the specified data with RAND Europe and the KSA research team. The governance of this data will be outlined in a Data Sharing Agreement between KSA and RAND Europe. **Schools will notify KSA immediately if a teacher or pupil wishes to withdraw from the intervention or the data collection**.

In addition, schools are asked to allow RAND Europe to match their pupil data to the National Pupil Database. No individual child, teacher or staff member will be identified in any reports arising from the research.

Monitoring of teacher concerns

Teacher surveys and Research Leader surveys will be used to monitor feedback from teachers on the development of the trial. The midline survey will be completed roughly three months after treatment assignment and constitutes the most important point in the evaluation at which concerns can be raised.

One month after the midline teacher and Research Leader surveys are completed, the evaluation team will send a report to the intervention team summarising the responses to the teacher and Research Leader surveys related to concerns raised about the intervention. While the evaluation team will not be able to identify the teachers that are experiencing negative consequences (because they will not know the identity of the responding teacher), they will be able to list the schools at which the concerns were raised.

The question that will be used to measure concerns among teachers and Research Leaders is:

'Are there any concerns over the use of digital feedback in your classroom, or more generally, that you would like to report?' [If 'yes', participants will be asked to specify in an open text box]

In addition, head teachers will monitor their own teachers as usual over the course of the year. Head teachers will be able to report any concerns they have regarding teachers in their existing interactions with the implementation team as part of their ongoing working relationship. The implementation team can inform the evaluation team of any concerns.

If there are any concerns it will first be necessary for the implementation team and the 'affected' schools to review the concern. It could be that additional training or information sessions are required. Should the concerns be of a more serious nature, the implementation team, the evaluation team and the school will review whether or not it is appropriate for the teacher to continue with digital feedback.

Special measure for schools with small numbers of participating teachers:

In schools where only small number of teachers participates a risk exists that teacher confidentiality cannot be maintained in the reporting. For example, if only one teacher participates, the anonymity of this teacher cannot be guaranteed. Combined reporting will therefore be used for schools with small numbers of participating teachers.

This does mean that if one of the teachers raises concerns, all of the schools in this combined group will be asked to take the steps outlined above.

6. The role of Kyra Schools Alliance implementation team

The implementation team will develop the intervention in partnership with Bishop Grosseteste University, and help monitor implementation. The implementation team will also work with RAND Europe to advise RAND Europe on the evaluation, such as the surveys.

KSA role and responsibilities for the intervention

- **Recruitment** - recruitment of schools and teachers within the constraints placed on the trial and reporting of recruitment progress.
- **Information** - provision of initial trial background information to participating schools and teachers and provision of a support to teachers via the email address research@unitedlearning.org.uk.
- **Training** – provision of initial training to teachers and research leads and facilitation of monthly professional learning communities.
- **Provision of the app** to all participating schools in the treatment condition.

KSA role and responsibilities for the evaluation

- **Consent** – distribution and collection of teacher and pupil consent forms in association with participating schools.
- **Data collection** - work with schools to collect data required for the evaluation and outlined in detail in the Data Sharing Agreements between RAND Europe and KSA and provision of that data to RAND Europe.
- **Keeping of records** - collation of a list of schools and pupils included in the trial that allow for later secure data linkage; creation and maintenance of a key contact list for all schools during the evaluation; reporting on attendance of teachers and RLs at initial training sessions; monitoring of attendance of teachers at PLC meetings; reporting on schools or teachers who withdraw from the evaluation; and sharing of the above records with the evaluation team.
- **Tracking of RL motivation** – providing a rating for each RL recruited against set criteria agreed with the evaluation team.
- **Reporting on trial progress** - reporting on the implementation of the Digital Feedback treatment.
- **Facilitating access to schools** for evaluation team fieldwork.
- **Distributing links to online surveys** to teachers, RLs and pupils.

7. The role of the independent evaluation team (RAND Europe)

The independent evaluation of the Digital Feedback intervention will be conducted by RAND Europe. The evaluation team will oversee data collection and will produce the evaluation report required by the Education Endowment Foundation.

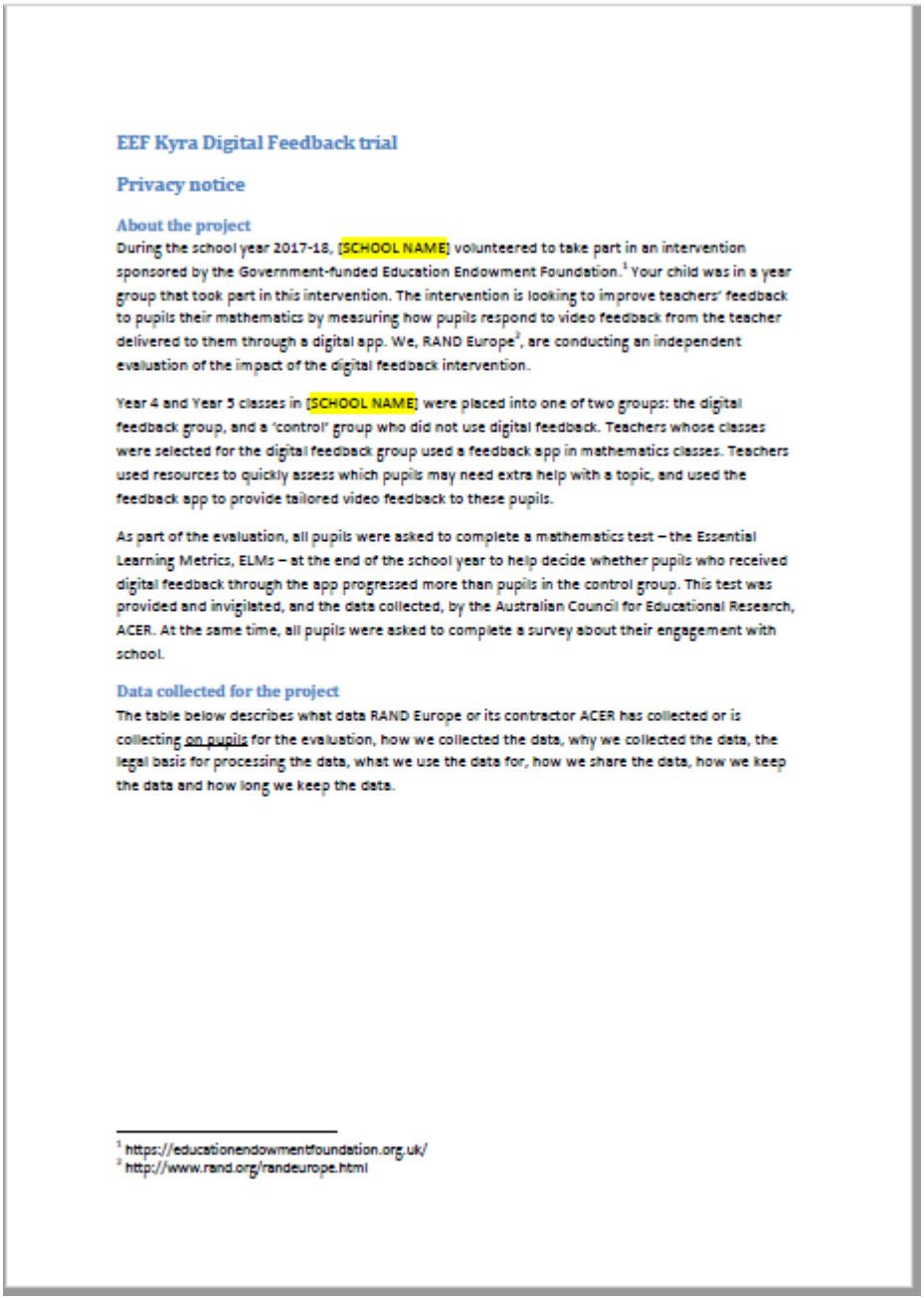
RAND Europe role and responsibilities for the evaluation

RAND Europe will design and conduct the independent evaluation. This includes:

- **Randomisation of participating year groups** to treatment/control conditions.
- **Development and implementation of data collection tools** -
    - Surveys
        - Teacher surveys: to be administered at three points over the year.
        - RL survey: to be administered at three points over the year.
        - Head teacher survey: to be administered at three points over the year.
        - Pupil engagement test: to be administered upon completion of the intervention.
    - Telephone interviews with a small selection of teachers and RLs following completion of the Trial.
    - Review of a sample of recorded digital feedback.

  - Review of a sample of diagnostic tests.
  - Review of a sample of RL coaching records.
- **Organisation of external mathematics testing** of all pupils in participating classes (both treatment and control). The test will be implemented by a third party.
- **Data storage** – Receipt and secure storage of all Trial data in line with the Data Sharing Agreements between RAND Europe and KSA.
- **Analysis of all Trial data.**
- **Production of the final evaluation report.**

Appendix D: Privacy notice for schools



EEF Kyra Digital Feedback trial

Privacy notice

About the project

During the school year 2017-18, [SCHOOL NAME] volunteered to take part in an intervention sponsored by the Government-funded Education Endowment Foundation.[1] Your child was in a year group that took part in this intervention. The intervention is looking to improve teachers' feedback to pupils their mathematics by measuring how pupils respond to video feedback from the teacher delivered to them through a digital app. We, RAND Europe[2], are conducting an independent evaluation of the impact of the digital feedback intervention.

Year 4 and Year 5 classes in [SCHOOL NAME] were placed into one of two groups: the digital feedback group, and a 'control' group who did not use digital feedback. Teachers whose classes were selected for the digital feedback group used a feedback app in mathematics classes. Teachers used resources to quickly assess which pupils may need extra help with a topic, and used the feedback app to provide tailored video feedback to these pupils.

As part of the evaluation, all pupils were asked to complete a mathematics test – the Essential Learning Metrics, ELMs – at the end of the school year to help decide whether pupils who received digital feedback through the app progressed more than pupils in the control group. This test was provided and invigilated, and the data collected, by the Australian Council for Educational Research, ACER. At the same time, all pupils were asked to complete a survey about their engagement with school.

Data collected for the project

The table below describes what data RAND Europe or its contractor ACER has collected or is collecting on pupils for the evaluation, how we collected the data, why we collected the data, the legal basis for processing the data, what we use the data for, how we share the data, how we keep the data and how long we keep the data.

[1] https://educationendowmentfoundation.org.uk/
[2] http://www.rand.org/randeurope.html

What data was collected	How we collected the data	Why we collected the data	Legal basis for processing	What we use the data for	How we share the data	How we keep the data secure	How long we keep the data
Control and Intervention groups Pupil data: name, UPN, year group, date of birth, gender, pupil premium, FSM status, home language, SEN provision, entry date, Key Stage 1 result in Math	We collected this data from schools at two time points: (1) in summer 2017, pupil data (except pupil name) was collected by the Kyra Teaching School Alliance on our behalf; (2) in spring 2018, the data was collected by ACER on our behalf. Data was collected at two time points as pupil information may have changed during that time.	We collected the data to be able to understand the characteristics of pupils who took part in the trial and to enable us to analyse whether the digital feedback intervention helped pupils progress more than pupils who did not receive the intervention.	[SCHOOL NAME] agreed to participate in the digital feedback trial and RAND Europe was commissioned to undertake the evaluation of the trial. RAND Europe has collected this information in its legitimate interests to fulfil the contractual requirement of the EEF evaluation. The rights and freedoms of the pupils will not be affected by the processing. ACER in collecting pupil data and conducting testing does so to perform a task carried out in the public interest.	We are using pupil names, UPNs and dates of birth to request further data about pupils from the National Pupil Database. As the UPN is not always unique, using pupil names and dates of birth allows us to ensure that we are collecting data about the correct pupils. The remaining data will be used in our analyses of whether the digital feedback intervention helped pupils progress more than pupils who did not receive the intervention.	Anonymised pupil information will be passed for archiving to the Fischer Family Trust. The data was not transferred outside of the EEA.	The data from the Kyra Teaching School Alliance was shared with RAND Europe in a password-protected file. The data from ACER was shared with RAND Europe using a secure file transfer platform. The data is stored securely on RAND Europe's servers.	The data will be stored securely on RAND Europe's data servers for the duration of the Digital Feedback evaluation. To allow us time to analyse and report the results of the trial, this period will extend beyond [SCHOOL NAME's] participation in the trial.
Intervention group only [In six 'case study' schools] Pupils' diagnostic assessments completed in class. Diagnostic assessments may include the names of the pupils.	This data was collected by teachers administering the diagnostic assessments as part of the digital feedback trial. We asked teachers to digitalise (scan, or take photos of) this data.	We collected the data to help us evaluate whether digital feedback is an effective intervention and for us to produce a report to the Education Endowment Foundation as part of our evaluation.	[SCHOOL NAME] agreed to participate in the digital feedback trial and RAND Europe was commissioned to undertake the evaluation of the trial. RAND Europe has collected this information in its legitimate interests to fulfil the contractual requirement of the EEF evaluation. The rights and freedoms of the pupils will not be affected by the processing. ACER in collecting pupil data and conducting testing does so to perform a task carried out in the public interest.	We will code the diagnostic assessments to assess pupils' understanding of particular topics and evaluate the function the diagnostic tests perform in tailoring/targeting the digital feedback, especially delayed feedback, to those children with greater learning needs. Names included in diagnostic assessments are incidental and will be removed during coding.	The data will not be shared. The data was not transferred outside of the EEA.	Teachers shared this with us through a secure file sharing platform. The data is stored securely on RAND Europe's servers.	The data will be stored securely on RAND Europe's data servers for the duration of the Digital Feedback evaluation. To allow us time to analyse and report the results of the trial, this period will extend beyond [SCHOOL NAME's] participation in the trial.
Intervention group only [In six 'case	This data was generated by teachers filming digital feedback	We collected the data to help us evaluate whether digital	[SCHOOL NAME] agreed to participate in the digital feedback	We will code the feedback videos to understand how teachers	The data will not be shared. The data was not transferred	We asked teachers to share the videos with us through	The data will be stored securely on RAND Europe's data

What data was collected	How we collected the data	Why we collected the data	Legal basis for processing	What we use the data for	How we share the data	How we keep the data secure	How long we keep the data
study' schools] Feedback videos prepared by teachers for pupils. Feedback videos mention the pupil name	videos for pupils as part of the digital feedback trial.	feedback is an effective intervention and for us to produce a report to the Education Endowment Foundation as part of our evaluation.	trial and RAND Europe was commissioned to undertake the evaluation of the trial. RAND Europe has collected this information in its legitimate interests to fulfil the contractual requirement of the EEF evaluation. The rights and freedoms of the pupils will not be affected by the processing. ACER in collecting pupil data and conducting testing does so to perform a task carried out in the public interest.	apply principles of good feedback in providing feedback to pupils. No pupil names or other personal data will be recorded in the coding.	outside of the EEA.	a secure file sharing platform. The data is stored securely on RAND Europe's servers.	servers for the duration of the Digital Feedback evaluation. To allow us time to analyse and report the results of the trial, this period will extend beyond [SCHOOL NAME's] participation in the trial.
Control and Intervention groups ELMs data and pupil	This data was collected by ACER as part of their administering and invigilating	We collected this data to help be able to test whether pupils who received digital feedback	[SCHOOL NAME] agreed to participate in the digital feedback trial and RAND Europe was	We will analyse the ELMs test data to evaluate whether digital feedback provided to	Anonymised pupil information will be passed for archiving to the Fischer Family	The data will be shared with RAND Europe using secure file transfer platforms. The	The data will be stored securely on RAND Europe's data servers for the duration of the

What data was collected	How we collected the data	Why we collected the data	Legal basis for processing	What we use the data for	How we share the data	How we keep the data secure	How long we keep the data
engagement data, including pupil names	of the ELMs mathematics test at the end of the school year. ACER collected the paper copies of the ELMs test and entered the data into an electronic database.	through the app progressed more than pupils in the control group, and for us to produce a report to the Education Endowment Foundation as part of our evaluation.	commissioned to undertake the evaluation of the trial. RAND Europe has collected this information in its legitimate interests to fulfil the contractual requirement of the EEF evaluation. The rights and freedoms of the pupils will not be affected by the processing. ACER in collecting pupil data and conducting testing does so to perform a task carried out in the public interest.	students is an effective intervention. We will compare the test scores to determine differences between pupils that have been provided digital feedback with those that have not. Pupil names will be used to match pupils' test scores to their characteristics ('pupil data', above). This will enable us to factor in pupils' characteristics when evaluating whether pupils who received digital feedback through the app progressed more than pupils in the control group.	Trust. The data was not transferred outside of the EEA.	data will be stored securely on RAND Europe's servers.	Digital Feedback evaluation. To allow us time to analyse and report the results of the trial, this period will extend beyond [SCHOOL NAME's] participation in the trial.

Your rights and choices

You may have certain rights in respect to processing of your data. These rights are as follows:

- The right of access to personal data that RAND Europe or ACER holds about a pupil
- The right to rectification of personal data and test results that RAND Europe or ACER holds about a pupil
- The right to erasure of your data prior to conclusion of the evaluation
- The right to restrict processing of pupil scores in the study
- The right to data portability
- The right to object to any processing that is described in this notice

To exercise any of these rights, please contact the RAND Europe data protection officer (redpo@rand.org).

How do you contact us

If you have any questions about how your data will be used, please do not hesitate to contact the RAND Europe data protection officer (redpo@rand.org). You may also contact the UK Information Commissioner's Office if you have any concerns about our use of your data at https://ico.org.uk/concerns/.

Appendix E: IPE details

Analysis of feedback videos

Here we provide further details on the coding framework used to analyse feedback videos. Reviewers looked at whether the following elements were present in the feedback recorded in the videos:

- Positive elements: 1 - Feedback draws attention to positive elements of the performance, for example, the details of correct responses; or 0 - Feedback is focused solely on incorrect responses.
- Constructive advice: 1 - Feedback includes constructive criticism: advice that provokes the student to improve task performance; or 0 - Feedback that does not provide information or support to improve performance or understanding.
- Changes in performance: 1 - Feedback refers to changes in performance from previous efforts; or 0 - There is a focus on comparisons with other students, or marks and grades.
- Self-assessment: 1 - Feedback includes an element of self-assessment by students (including peer assessment) as part of the process of encouraging student autonomy and responsibility; or 0 - A reliance on extrinsic rewards (stickers, stars) or feedback includes punishment.
- Task-specific: 1 - Feedback provides information about a task, how well it was performed and how to do it more effectively; or 0 - Non-specific feedback is given: e.g. praise or criticism for task performance without detail.
- Process-specific: 1 - Feedback at the process level: how can the student improve the learning processes needed to understand and perform the task?; or 0 - Feedback at the self-level: comment on personal qualities of the student, either positive or negative that provides little or no information about processes or performance.
- Metacognitive feedback: 1 - Feedback at the self-regulation level: how can the student do a better job of planning, monitoring and managing their actions and using strategies in approaching the task? This is also described as "metacognitive" feedback; or 0 - Feedback at the self-level: comment on personal qualities of the student, either positive or negative that provides little or no information about processes or performance.

Coding protocol used

Score 1- Effective feedback
Score 0- Ineffective feedback
N/A - not applicable

Video ID	Kind of feedback				Level of feedback			
	Positive elements	**Constructive advice**	**Changes in performance**	**Self-assessment**	**Task-specific**	**Process-specific**	**Metacognitive feedback**	**Comments**
…								
…								
…								
…								
Total no. of videos with element								

Analysis of coaching records

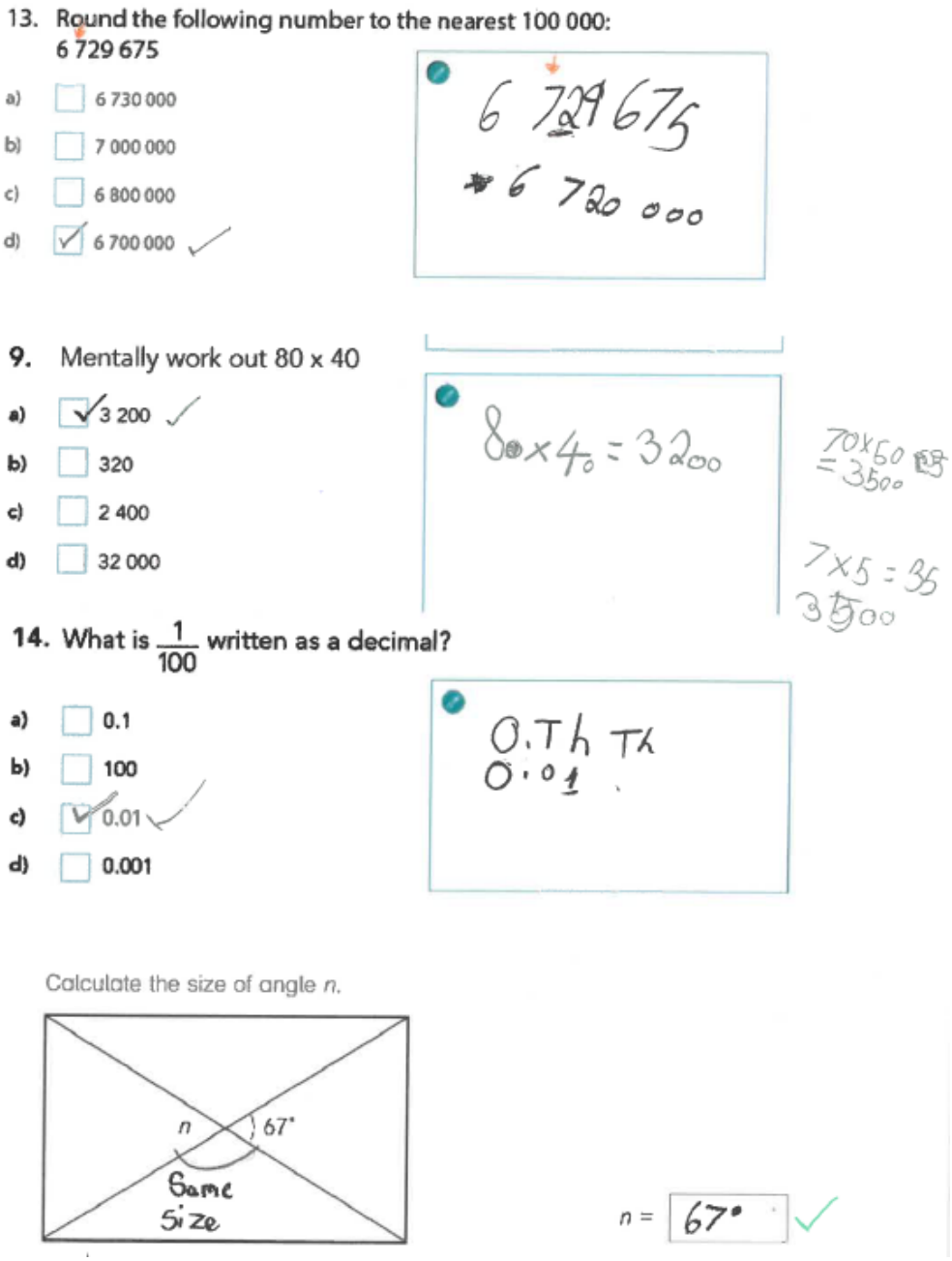
Below we include the coding protocol used to review the coaching records that the case study schools shared with the evaluation team.

Coaching record coding protocol

Category	School		
Evidence of/related to:	**School name 1 (Teacher: name RL: Name)**	**School name 2 (Teacher: name RL: Name)**	**School name 3 (Teacher: name RL: Name)**
Is a range of topics touched upon? What are these? Please provide examples.	[Yes/no/additional comments]		
○ After meeting action plan included? (Goal - G of GROW)			
○ Support provided by RL? In what form? Is this reflected in the coaching records?			
○ Coach helps define how to re-teach a concept more effectively			
○ Think about 'the questions' the teacher should be asking			
○ Think about how to breakdown a task – so teacher can improve their teaching			
○ Evidence on whether RLs/teachers are 'tracking' developments/improvements in how teachers give feedback (R - review in GROW)			
Descriptives			
No. of records for teacher			
Length of meetings			
Frequency of the meetings			

Example of diagnostic assessment

The example below was provided by case study school A.



For the intervention, diagnostic assessments were developed by the Kyra Teaching Alliance and made available to schools.

Appendix F: Siddle (2015) poster

St Margaret's CE Primary School, Withern, Lincolnshire

Verbal and visual-digital feedback on creative writing improves progress rates compared to written feedback in rural primary schools – a preliminary study

James Siddle - Kyra Teaching School Alliance
james.siddle@st-margerst-pri.lincs.sch.uk

Introduction

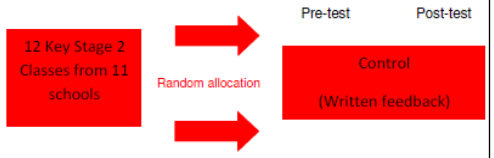
Research evidence suggests that effective feedback has a significant impact on pupil progress. Our work over the last 18 months across 34 schools has indicated that delivering effective feedback practices can be challenging and specific approaches are required to improve pupil outcomes. Trials across collaborative partnerships of schools have indicated a positive correlation regarding the impact of digital feedback on outcomes in writing. Evidence suggests greater pupil engagement with feedback and improved outcomes in response to the feedback. Our research also suggests the impact of digital feedback may be greatest on lower attaining and free school meal (FSM) children.

This is an important area to explore using a randomised controlled trial design because the initial small scale experimental trial we have undertaken on digital feedback suggests these techniques can have a positive effect on pupils' writing outcomes, but variation between schools indicates the approach needs further refinement. It is an approach (using digital technology) that is poorly studied at a time when many schools are investing significantly in new digital technology. Even for schools without tablet technology verbal and visual feedback can be given via video. Research suggests that marking times may also be reduced using this form of feedback.

Research Design

A between-subject design was used with a pre and post-test. To address the aims of the research the independent variable of intervention type was operationalised by creating two conditions:

- IV 1 (Control condition) – written feedback
- IV 1 (Control condition) – written feedback



Method: Participants

Twelve classes from eleven rural school primary schools participated in the study. Pupils were able to be randomly allocated to a control and intervention group in each class. In total, 231 Key Stage 2 pupils (120 boys and 111 girls) took part in the study (113 in control and 118 in the intervention). The total of FSM pupils was 42 or 18.18% which is below the national average (NA) of 26.6%. The total of SEND pupils was 40 pupils or 17.3% which is slightly above NA which is 16.6%.

Procedure

The randomly allocated groups were both given a writing prompt, success criteria rubric and a standard video input delivered together with a short film as a writing stimulus. The pupils then had ten minutes planning time and 40 minutes writing time. The control group then received written feedback; the intervention group feedback digitally. Each group then had the same amount of 'fix it' time the following day. Pupils made corrections in the same format (written form); pupils also recorded 'what I have learnt' statements, either in written form or digitally (depending whether they were control or intervention). All pupils were then given another piece of creative writing (using the same genre) the following day. The procedure was repeated and rubrics/models used. The work was then marked against the two success criteria points given at the feedback stage and the gain scores recorded.

Materials

A rubric using success criteria for all pupils to use at the start of the trial; a model for use by all pupils was used; A standardised way of delivering written feedback was introduced and models of giving digital feedback via an ipad shared with class teachers to standardise this approach (through video); A format given to pupils regarding how to correct their work following feedback

Conclusions:

- The gains, although needing further research (due to small sub-group samples), match current research evidence around the impact of digital technology on closing the gap in attainment , which suggest digital technology may produce gains of +4 months; (https://educationendowmentfoundation.org.uk/index.php/toolkit/toolkit-a-z/digital-technology/) ;
- The data suggests that the intervention produces the greatest gains for disadvantaged and SEND pupils and may be an effective intervention for closing the gap;
- Surveyed perceptions of digital feedback indicates pupils , in general, feel they make better progress in their written work following digital feedback and this backs up the findings of this RCT.

Suggestions for Future Research:

- Our earlier research suggests the gains may be even more substantial in maths and it would be productive to research the effect in different subject areas;
- The results for SEND and FSM pupils show promise but the trial would require replication with greater numbers to produce a more secure indication of impact;
- The research suggests that boys may make greater gains using this intervention and further research could help clarify if and why this is the case;
- The effect of the intervention was greatest on SEND pupils. However, our data does not take into account the needs of these pupils; it would be useful to identify the effect on different groups of SEND.

Results

Disadvantaged Pupils (n=43; Intervention (I) =21 Control (C) =22)

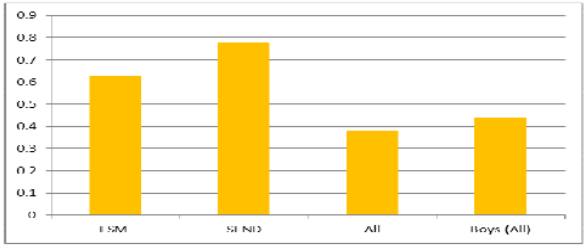
Gain scores were not normally distributed. A Mann-Whitney U test was applied. This showed a impact on progress for the intervention (verbal and visual feedback) compared to the control (written feedback) r = 0.308 and d=0.63 (p = 0.03 (one tailed)); **this indicates +8 months gain - if r is converted to d and the EEF convention is applied.**

SEND Pupils (n=40; I =20 C =20)

Again, a Mann-Whitney U test was applied to gain scores. This showed a significant impact on progress for pupils who experienced the intervention r = 0.37 and d=0.78 (p = 0.013 (one tailed)); **+9 months gain**

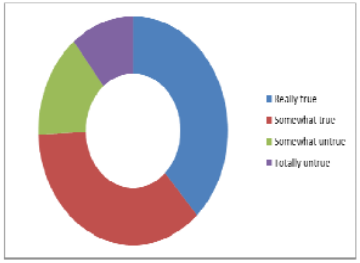
All Pupils (n=231; I =118 C =113; Boys n= 120; I=58 C=62)

Finally, a Mann-Whitney U test assessed the data for all pupils. This again showed that the intervention impacted on progress, but with a smaller effect size r = 0.186 and d= 0.38 (p = 0.004 (one tailed)); **+5 months gain**. However, the results for all boys were slightly higher r = 0.218 and d= 0.44 (p = 0.012 (one tailed)) indicating a greater positive impact for boys with **+5 months gain**.

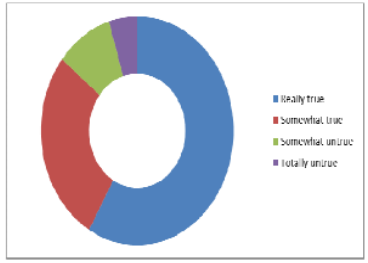


Pupil perceptions of written vs digital feedback (n=153)

I think my work gets better following written feedback from my teacher



I think my work gets better following digital feedback via an ipad from my teacher



Appendix G: Data sharing schedule for Case Studies

SCHEDULE 1
SINGLE POINT OF CONTACT

(a) Julie Belanger, Principal Investigator, Digital Feedback in Primary Maths evaluation, RAND Europe, EEF_DF@rand.org

(b) [Party 1's SPoC – name, role, team and contact details]

SCHEDULE 2
PURPOSE To allow RAND Europe to evaluate the effectiveness of Digital Feedback as commissioned by the Education Endowment Foundation.

SCHEDULE 3
DATA

Item No	Description of Data	Personal Identifiable Data (Yes/no/special category)
1	Pupil name if associated with feedback videos for pupils that are recorded by teachers	Yes
2	Teacher names in coaching records between Research Leads and teachers	Yes
3	Pupil names in diagnostic assessments of pupils	Yes

SCHEDULE 4
FURTHER DETAIL ON SHARED PERSONAL DATA AND ACCESS AND PROCESSING RESTRICTIONS

Item No	1
Discloser	Party 1
Receiver	RAND Europe
Purpose of processing	To permit RAND Europe to evaluate whether digital feedback provided to students is an effective intervention and for RAND Europe to produce a report to the Education Endowment Foundation on this subject
Legal basis	Legitimate Interest to undertake the Study commissioned by the Education Endowment Foundation to look at maths ability and pupil engagement. The rights and freedoms of the children receiving the feedback will not be affected as the feedback videos will not be deliberately associated with names. When watched the names will not be recorded and the videos will ultimately be deleted. Therefore no decisions will be made about the children on the basis of this information.
Nature of processing	Immediate and delayed feedback videos will be evaluated and coded to understand how teachers apply principles of good

	feedback in providing feedback to pupils. No personal data will be recorded in the coding.
Permitted Recipients	**None**
Data minimisation	**Any names of pupils or pupils will not be included in the analysis.**
Access and processing restrictions?	**Syncplicity will be used to transfer the files from schools. RAND will only process the video files for the purposes of the Project and the videos will be deleted and data destroyed at that stage.**
Deletion Procedure	**RAND to delete videos after coding.**

Item No	**2**
Discloser	**Party 1**
Receiver	**RAND Europe**
Purpose of processing	**To permit RAND Europe to evaluate whether digital feedback provided to students is an effective intervention and for RAND Europe to produce a report to the Education Endowment Foundation on this subject.**
Legal basis	**Legitimate Interest to undertake the Study commissioned by the Education Endowment Foundation to look at maths ability and pupil engagement. The rights and freedoms of the teachers will not be affected as their coaching records will be coded and names removed**
Nature of processing	**Coaching records between Research Leads and teachers will be evaluated to understand how the intervention is being carried out in schools and coded records will be created that remove all personal data.**
Permitted Recipients	**None**
Data minimisation	**Names of teachers and research leads will be removed during analysis so that the data are not identifiable. Schools will remain identified so small numbers (for research leads and some teachers) may enable identification. No resulting report will identify schools in association with the data.**
Access and processing restrictions?	**RAND Europe is only entitled to pass data to ACER for the testing and then to retain data until tests are provided by ACER and analysed by RAND Europe.**
Deletion Procedure	**Will be maintained for duration of project and thereafter for academic research purposes.**

Item No	**3**
Discloser	**Party 1**
Receiver	**RAND Europe**
Purpose of processing	**To permit RAND Europe to evaluate whether digital feedback provided to students is an effective intervention and for RAND Europe to produce a report to the Education Endowment Foundation on this subject**
Legal basis	**Legitimate Interest to undertake the Study commissioned by the Education Endowment Foundation to look at maths ability and pupil engagement. The rights and freedoms of the children undertaking the diagnostic assessments will not be affected as their names will be dissociated from the coded database.**
Nature of processing	**Names included in diagnostic assessments are incidental and will be reviewed through coding (that removes the names) to asses pupils' understanding of particular topics and evaluate the function the diagnostic tests perform in tailoring/targeting the digital feedback, especially delayed feedback, to those children with greater learning needs.**

Permitted Recipients	**None**
Data minimisation	**Names of pupils are unlikely to be associated with the diagnostic assessments as the data will be taken from material that does not record pupil names. Where loose work sheets or names are incidental these will not be referred to in later coded results.**
Access and processing restrictions?	**Diagnostic will be coded and then the original assessments will be deleted. The coded database will not include personal data.**
Deletion Procedure	**To be deleted by RAND Europe once coding complete.**

SCHEDULE 5
APPROPRIATE TECHNICAL AND ORGANISATIONAL SECURITY MEASURES

Signed by [NAME OF DIRECTOR] for and on behalf of [NAME OF DATA DISCLOSER]

.....

Director

Signed by [NAME OF DIRECTOR] for and on behalf of [NAME OF DATA RECEIVER]

.....

Director

Appendix H: Details on data shared between Kyra Teaching Alliance and RAND Europe

- Teachers:
    - Email address teacher
    - First name
    - Middle name(s) (if any)
    - Surname (maiden name if married)
    - School name
    - Year group (2016/17)
    - Year group (2017/18)
    - Age
    - Gender
    - Length of service
    - Full time/part time contract
    - Highest degree earned (BA, MSc, PhD)
    - Certification
- Research leads:
    - First name
    - Middle name(s) (if any)
    - Surname (maiden name if married)
    - School name
    - Age
    - Gender
    - Length of service
    - Full time/part time contract
    - Highest degree earned (BA, MSc, PhD)
- Pupils:
    - UPN (Unique Pupil Number)
    - First name
    - Surname
    - School name
    - Teacher name
    - Year group (2016/17)
    - Year group (2017/18)
    - Date Of Birth
    - Gender
    - Pupil Premium
    - Home Language
    - Entry Date
    - SEN Provision (Special Education Needs)
    - FSM (Free School Meal)
    - Key Stage 1 results (prior attainment)

Finally, for pupils, RAND will also receive the names of pupils who through the parental opt-out forms have opted out from the evaluation.

Appendix I: Impact evaluation analysis graphs and tables

OUTCOME MEASURES DISTRIBUTIONS

Figure 19: ELM raw score distribution

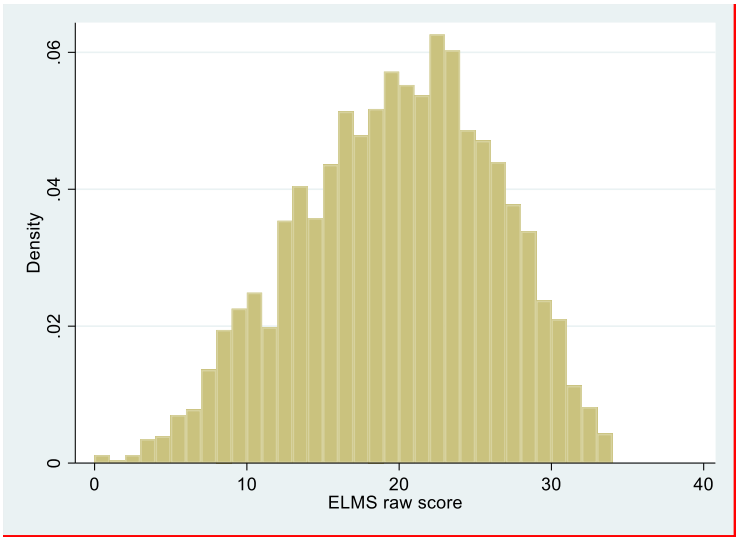
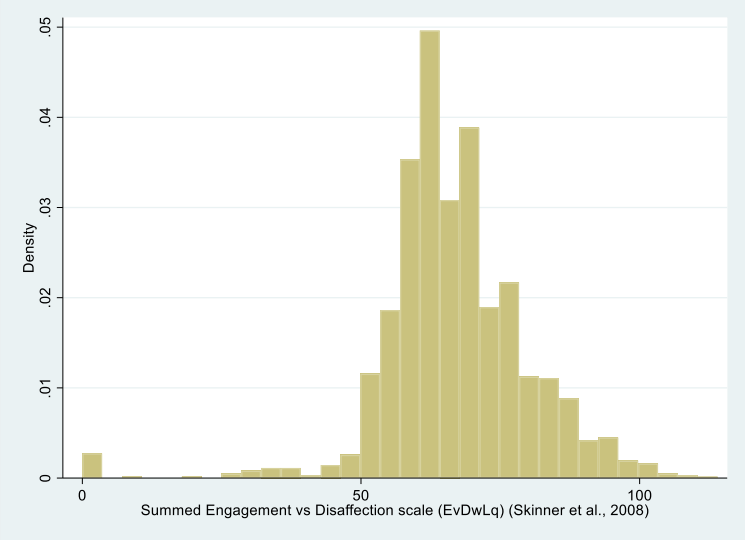


Figure 20: Disengagement measure distribution (EvDwLq)



Impact evaluation analysis code example

ITT model for Primary outcome (ELM mathematics)

```
mixed raw i.ks1baseline_r treatment year1718 || mlm_id:, mle vce(robust)
```

ITT model for FSM sub-group analysis (ELM mathematics)

```
mixed raw i.ks1baseline_r treatment year1718 everfsm_6 fsm_t || mlm_id:,
mle vce(robust)
```

ITT model for secondary outcome (disengagement from school)

```
mixed engagement_sum treatment i.ks1baseline_r year1718 || mlm_id:, mle
vce(robust)
```

ITT model for FSM sub-group analysis (disengagement from school)

```
mixed engagement_sum treatment i.ks1baseline_r year1718 everfsm_6 fsm_t ||
mlm_id:, mle vce(robust)
```

MAIN ANALYSIS DETAILS

Table 14: ELM mathematics ITT analysis results (n=2,133 pupils)

ELM mathematics	Coeff.	se	z	p	95% CI	95% CI
Treatment (0=control 1=treatment)	-0.29	0.43	-0.66	0.51	-1.13	0.56
KS1 – below standard (ref)	n/a	n/a	n/a	n/a	n/a	n/a
KS1 – At/above standard	7.99	0.35	22.96	0.00	7.31	8.67
KS1 – Missing	4.30	1.09	3.96	0.00	2.17	6.43
Year group (Yr5 vs Yr4)	1.27	0.38	3.37	0.00	0.53	2.01
Intercept	6.96	1.75	3.98	0.00	3.53	10.38
Random effects						
var(_intercept)	2.66	0.85			1.43	4.96
var(Residual)	28.30	1.23			25.99	30.81
ICC (between class)	0.09	0.03			0.05	0.15

Table 15: Disengagement survey ITT results (n=1,572)

EvDwLq	Coef.	se	z	p	95% CI	95% CI
Treatment (0=control 1=treatment)	-1.18	0.86	-1.37	0.17	-2.87	0.51
KS1 – below standard (ref)	n/a	n/a	n/a	n/a	n/a	n/a
KS1 – At/above standard	-3.45	0.95	-3.64	0.00	-5.31	-1.60
KS1 – Missing	-4.14	2.04	-2.03	0.04	-8.13	-0.14
Year group (Yr5 vs Yr4)	2.62	0.85	3.09	0.00	0.96	4.28
Intercept	58.52	4.07	14.38	0.00	50.54	66.49
Random effects						
var(_intercept)	5.73	1.99			2.90	11.30
var(Residual)	161.17	16.35			132.11	196.62
ICC (between class)	0.03	0.01			0.02	0.06

FSM SUB-GROUP ANALYSIS

Table 16: ELM mathematics FSM sub-group analysis (n=2,133)

ELM mathematics	Coeff.	se	z	p	95% CI	95% CI
KS1 – below standard (ref)	n/a	n/a	n/a	n/a	n/a	n/a
KS1 – At/above standard	7.79	0.36	21.94	0.00	7.10	8.49
Year group (Yr5 vs Yr4)	-0.63	0.44	-1.43	0.15	-1.49	0.23
Ever6FSM	1.34	0.37	3.63	0.00	0.62	2.07
Treatment	-1.79	0.37	-4.84	0.00	-2.51	-1.06
FSM * treatment	0.92	0.49	1.87	0.06	-0.04	1.88
Intercept	7.42	1.69	4.38	0.00	4.10	10.74
Random effects						
var(_cons)	2.37	0.76			1.27	4.43
var(Residual)	27.59	1.22			25.31	30.08
ICC between-class	0.08	0.02			0.04	0.14

Table 17: Disengagement survey FSM sub-group analysis (n=1,512)

Disengagement	Coeff.	se	z	p	95% CI	95% CI
KS1 – below standard (ref)	n/a	n/a	n/a	n/a	n/a	n/a
KS1 – At/above standard	-3.26	1.02	-3.20	0.00	-5.25	-1.26
Year group (Yr5 vs Yr4)	2.58	0.89	2.89	0.00	0.83	4.34
Ever6FSM	2.34	1.11	2.11	0.04	0.16	4.52
Treatment	-0.86	0.98	-0.87	0.38	-2.78	1.07
FSM * treatment	-1.27	1.51	-0.84	0.40	-4.22	1.69
Intercept	57.79	4.23	13.67	0.00	49.50	66.08
Random effects						
var(_cons)	6.33	2.11			3.30	12.15
var(Residual)	159.11	16.95			129.13	196.06
ICC between-class	0.04	0.01			0.02	0.07

SPECIFICATION CHECK RESULTS

Table 18: ELMS specification check results, single level OLS (n=2,133)

ELM mathematics	Coeff.	se	z	p	95% CI	95% CI
Treatment (0=control 1=treatment)	-0.11	0.26	-0.41	0.68	-0.63	0.41
KS1 – below standard (ref)	n/a	n/a	n/a	n/a	n/a	n/a
KS1 – At/above standard	7.81	0.35	22.07	0.00	7.11	8.52
KS1 – Missing	4.26	1.08	3.94	0.00	2.12	6.40
Year group (Yr5 vs Yr4)	1.40	0.27	5.24	0.00	0.87	1.93
Intercept	3.26	1.20	2.73	0.01	0.89	5.63

Table note: school dummy variable results omitted

Table 19: Disengagement survey specification check results, single level OLS (n=1,578)

Disengagement	Coeff.	se	z	p	95% CI	95% CI
Treatment (0=control 1=treatment)	-1.28	0.58	-2.21	0.03	-2.43	-0.13
KS1 – below standard (ref)	n/a	n/a	n/a	n/a	n/a	n/a
KS1 – At/above standard	-3.48	0.99	-3.49	0.00	-5.45	-1.50
KS1 – Missing	-4.44	2.06	-2.15	0.03	-8.54	-0.35
Year group (Yr5 vs Yr4)	2.39	0.61	3.90	0.00	1.17	3.60
Intercept	59.22	2.76	21.44	0.00	53.74	64.71

Table note: school dummy variable results omitted

MISSING DATA

Table 20: ELMS outcome missing data details

PANEL A: MISSING DATA - ALL TO CONTROL (N=2,564)

ELM mathematics	Coeff.	se	z	p	95% CI	95% CI
Treatment (0=control 1=treatment)	-0.29	0.43	-0.66	0.51	-1.13	0.56
KS1 – below standard (ref)	n/a	n/a	n/a	n/a	n/a	n/a
KS1 – At/above standard	7.99	0.35	22.96	0.00	7.31	8.67
KS1 – Missing	4.30	1.09	3.96	0.00	2.17	6.43
Year group (Yr5 vs Yr4)	1.27	0.38	3.37	0.00	0.53	2.01
Intercept	6.96	1.75	3.98	0.00	3.53	10.38

MISSING DATA - ALL TO TREATMENT (N=2,564)

ELM mathematics	Coeff.	se	z	p	95% CI	95% CI
Treatment (0=control 1=treatment)	-0.33	0.41	-0.79	0.43	-1.13	0.48
KS1 – below standard (ref)	n/a	n/a	n/a	n/a	n/a	n/a
KS1 – At/above standard	8.11	0.33	24.96	0.00	7.48	8.75
KS1 – Missing	4.74	0.99	4.81	0.00	2.81	6.67
Year group (Yr5 vs Yr4)	1.56	0.40	3.88	0.00	0.77	2.35
Intercept	5.66	1.84	3.08	0.00	2.06	9.26

Table 21: Disengagement outcome missing data details

PANEL A: MISSING DATA - ALL TO CONTROL (N=1,742)

EvDwLq	Coeff.	se	z	p	95% CI	95% CI
Treatment (0=control 1=treatment)	-1.20	0.81	-1.47	0.14	-2.79	0.40
KS1 – below standard (ref)	n/a	n/a	n/a	n/a	n/a	n/a
KS1 – At/above standard	-3.33	0.95	-3.51	0.00	-5.20	-1.47
KS1 – Missing	-3.86	1.95	-1.98	0.05	-7.67	-0.05
Year group (Yr5 vs Yr4)	2.59	0.74	3.49	0.00	1.14	4.04
Intercept	58.57	3.53	16.61	0.00	51.66	65.48

PANEL B: MISSING DATA - ALL TO TREATMENT (N=1,742)

EvDwLq	Coeff.	se	z	p	95% CI	95% CI
Treatment (0=control 1=treatment)	-1.01	0.81	-1.26	0.21	-2.59	0.57
KS1 – below standard (ref)	n/a	n/a	n/a	n/a	n/a	n/a
KS1 – At/above standard	-3.32	0.95	-3.48	0.00	-5.19	-1.45
KS1 – Missing	-3.85	1.94	-1.98	0.05	-7.66	-0.04
Year group (Yr5 vs Yr4)	2.67	0.73	3.64	0.00	1.23	4.11
Intercept	58.18	3.51	16.57	0.00	51.30	65.07

Appendix J: Survey response counts

In this appendix we provide the response counts for survey questions discussed in the implementation and process evaluation section.

Figure 4 counts

Teachers midline treatment responses (Figure 4)

With regard to the one-day training session, please indicate the extent to which you agree or disagree with the following statements:

Midline Treatment Teacher:	Strongly agree	Agree	Disagree	Strongly disagree	Don't know
The information provided was **unclear**	0	5	13	1	1
The information provided was **useful**	2	17	1	0	0
The information provided was **detailed enough**	2	12	3	1	2
It was possible to ask questions specific to my class	2	16	1	0	1
The trainer provided useful answers to my questions	2	14	2	0	2

Research Leads midline responses (Figure 4)

Please indicate the extent to which you agree or disagree with the following statements about the **app**:

Midline Research Lead	**Strongly agree**	**Agree**	**Disagree**	**Strongly disagree**	**Don't know**
During the training session, the information provided about the app was **unclear**	3	1	7	0	0
During the training session, the information provided about the app was **useful**	0	7	2	1	1
During the training session, the information provided about the app was **detailed enough**	0	6	3	1	1
During the training session, it was possible to ask questions specific to my school	2	6	2	0	1
During the training session, the trainer provided useful answers to my questions about the app	2	4	1	0	3

Figure 5 counts

Research Leads midline response counts (Figure 5)

Please indicate the extent to which you agree or disagree with the following statements about **coaching:**

Midline Research Lead	**Strongly agree**	**Agree**	**Disagree**	**Strongly disagree**	**Don't know**
During the training session, the purpose of the coaching provided by Research Leads was made clear to me	1	6	2	0	0

The training improved my coaching skills	1	3	3	1	1
The training did not provide information on coaching that was new to me	1	4	4	0	0
The trainer provided useful answers to my questions about coaching	1	3	2	0	3

Figure 6 counts

Teacher response counts midline (Figure 6)

To what extent do you feel the training prepared you to use digital feedback in your class?

Midline Treatment Teacher:	**Very prepared**	**Somewhat prepared**	**Not very prepared**	**Not at all prepared**
	3	13	4	0

Research Lead response counts midline (Figure 6)

To what extent do you feel the training prepared you to support teachers using digital feedback in your school? (**7 did not answer question)**

Midline Treatment Research Lead:	**Very prepared**	**Somewhat prepared**	**Not very prepared**	**Not at all prepared**
	1	6	1	1

Figure 11 counts

Teacher response counts midline and endline (Figure 11)

How often do you meet with the Research Lead in your school?

Midline Treatment Teacher:	**Several times a week**	**Once a week**	**Once every two or three weeks**	**Once a month**	**Less than once a month**	**Never**
	0	6	5	2	4	5

Endline Treatment Teacher:	**Several times a week**	**Once a week**	**Once every two or three weeks**	**Once a month**	**Less than once a month**	**Never**
	1	3	3	1	5	6

Research Lead response counts midline and endline (Figure 11)

Midline Research Lead	**Several times a week**	**Once a week**	**Once every two or three weeks**	**Once a month**	**Less than once a month**	**Never**
	2	4	9	1	0	0

Endline Research Lead	**Several times a week**	**Once a week**	**Once every two or three weeks**	**Once a month**	**Less than once a month**	**Never**
	2	5	5	0	4	1

Figure 12, Figure 13 and Figure 14 counts

Teacher response counts midline and endline (Figure 12, Figure 13 and Figure 14)

How do you use the results from the diagnostic tests in your daily teaching practices? *Please tick all that apply.*

Midline Teacher	**To focus teaching on areas that the class as a whole seems to be struggling with**	**To track progress made by the class as a whole, over the course of the year**	**To target digital feedback at pupils with greater learning needs, as identified by the diagnostic tests**	**Teachers do not use diagnostic tests**
Yes	20	11	14	23
No	3	12	9	0

Endline Teacher	**To focus teaching on areas that the class as a whole seems to be struggling with**	**To track progress made by the class as a whole, over the course of the year**	**To target digital feedback at pupils with greater learning needs, as identified by the diagnostic tests**	**Teachers do not use diagnostic tests**
Yes	16	11	10	1
No	3	8	9	18

Research Lead response counts midline and endline (Figure 12, Figure 13 and Figure 14)

How do teachers in your school who use digital feedback generally use the results from the diagnostic tests in their daily teaching practices? *Please tick all that apply.*

Midline Research Leads	**To focus teaching on areas that the class as a whole seems to be struggling with**	**To track progress made by the class as a whole, over the course of the year**	**To target digital feedback at pupils with greater learning needs, as identified by the diagnostic tests**	**Teachers do not use diagnostic tests**
Yes	13	7	9	1
No	3	9	7	15

Endline Research Leads	**To focus teaching on areas that the class as a whole seems to be struggling with**	**To track progress made by the class as a whole, over the course of the year**	**To target digital feedback at pupils with greater learning needs, as identified by the diagnostic tests**	**Teachers do not use diagnostic tests**
Yes	10	5	9	2
No	7	12	8	15

Head Teacher response counts midline and endline (Figure 12, Figure 13 and Figure 14)

How do teachers in your school who use digital feedback generally use the results from the diagnostic tests in their daily teaching practices?

Midline Head Teachers	**To focus teaching on areas that the class as a whole seems to be struggling with**	**To track progress made by the class as a whole, over the course of the year**	**To target digital feedback at pupils with greater learning needs, as identified by the diagnostic tests**	**Teachers do not use diagnostic tests**
Yes	4	3	7	8
No	4	5	1	0

Endline Head Teachers	**To focus teaching on areas that the class as a whole seems to be struggling with**	**To track progress made by the class as a whole, over the course of the year**	**To target digital feedback at pupils with greater learning needs, as identified by the diagnostic tests**	**Teachers do not use diagnostic tests**
Yes	5	3	4	0
No	2	4	3	7

Figure 7 and Figure 8 counts

Teacher response counts midline and endline (Figure 7 and 8)

How often do you use the following two types of digital feedback:

Midline Treatment Teacher:	**Once a day or more than once a day**	**Several times per week**	**Once a week**	**Once every two or three weeks**	**Once a month**	**Less than once a month**
Immediate digital feedback	0	1	5	5	3	9
Delayed digital feedback	0	6	6	7	1	3

Endline Treatment Teacher:	**Once a day or more than once a day**	**Several times per week**	**Once a week**	**Once every two or three weeks**	**Once a month**	**Less than once a month**
Immediate digital feedback	0	0	2	5	1	11
Delayed digital feedback	0	0	5	5	2	7

Research Lead response counts midline and endline (Figure 7 and 8)

How often do teachers in your school who use digital feedback use the following two types of digital feedback?

Midline Research Lead	**Once a day or**	**Several times per week**	**Once a week**	**Once every two**	**Once a month**	**Less than once a month**

	more than once a day			or three weeks		
Immediate digital feedback	0	4	3	7	0	2
Delayed digital feedback	0	2	7	4	0	3

Endline Research Lead	**Once a day or more than once a day**	**Several times per week**	**Once a week**	**Once every two or three weeks**	**Once a month**	**Less than once a month**
Immediate digital feedback	0	3	0	4	1	9
Delayed digital feedback	0	1	4	6	1	5

Figure 9 counts

Teacher response counts midline and endline (Figure 9)

With regard to the use of the app required for providing digital feedback, to what extent do you agree or disagree with the following statements

Midline Treatment Teacher:	**Strongly agree**	**Agree**	**Disagree**	**Strongly disagree**	**Don't know**
It is difficult to provide digital feedback using the app	4	5	11	2	1
Providing feedback with the app takes less time than providing similar written comments	0	6	9	4	4
It is easy to integrate digital feedback in daily teaching practices using the app	0	8	11	1	3

Endline Treatment Teacher :	**Strongly agree**	**Agree**	**Disagree**	**Strongly disagree**	**Don't know**
It is difficult to provide digital feedback using the app	4	7	5	0	3
Providing feedback with the app takes less time than providing similar written comments	1	1	9	4	4
It is easy to integrate digital feedback in daily teaching practices using the app	0	1	14	2	2

Figure 10, Figure 17 and Figure 18 counts

Teacher response counts midline and endline (Figure 10, Figure 17, Figure 18)

To what extent do you agree or disagree with the following statements on digital feedback (i.e. feedback given using digital technologies, either immediately or after class):

Midline Treatment Teacher:	**Strongly agree**	**Agree**	**Disagree**	**Strongly disagree**	**Don't know**

Digital feedback allows teachers to give more detailed and specific feedback than traditional written feedback	4	14	5	0	0
Pupils engage more with digital feedback than with traditional written feedback	3	16	1	0	3
Delayed digital feedback is useful to target feedback at specific pupils	3	15	3	0	2
Digital feedback is useful for specific subjects	2	17	1	0	3
During a class it is difficult to fit in immediate digital feedback	11	9	2	0	1
Pupils get distracted by digital feedback and do not pay attention to the content	1	8	11	1	2

Endline Treatment Teacher:	**Strongly agree**	**Agree**	**Disagree**	**Strongly disagree**	**Don't know**
Digital feedback allows teachers to give more detailed and specific feedback than traditional written feedback	1	12	4	0	2
Pupils engage more with digital feedback than with traditional written feedback	0	11	3	0	5
Delayed digital feedback is useful to target feedback at specific pupils	2	13	2	0	2
Digital feedback is useful for specific subjects	0	17	0	0	2
During a class it is difficult to fit in immediate digital feedback	9	9	0	0	1
Pupils get distracted by digital feedback and do not pay attention to the content	1	10	5	0	3

Research Lead response counts midline and endline (Figure 10, Figure 17, Figure 18)

To what extent do you agree or disagree with the following statements on digital feedback?

Midline Research Lead	**Strongly agree**	**Agree**	**Disagree**	**Strongly disagree**	**Don't know**
Digital feedback allows teachers to give more detailed and specific feedback than traditional written feedback	5	9	0	0	2
Pupils engage more with digital feedback than with traditional written feedback	3	9	2	1	1
Delayed digital feedback is useful to target feedback at specific pupils	1	12	2	0	1
Digital feedback is useful for maths	4	10	0	0	2
During a class it is difficult to fit in immediate digital feedback	8	5	3	0	0
Pupils get distracted by digital feedback and do not pay attention to the content	0	3	8	2	3

Endline	**Strongly agree**	**Agree**	**Disagree**	**Strongly disagree**	**Don't know**
Digital feedback allows teachers to give more detailed and specific feedback than traditional written feedback	3	8	1	0	5

Pupils engage more with digital feedback than with traditional written feedback	0	9	3	0	5
Delayed digital feedback is useful to target feedback at specific pupils	2	6	1	0	8
Digital feedback is useful for maths	2	12	2	0	1
During a class it is difficult to fit in immediate digital feedback	10	4	1	0	2
Pupils get distracted by digital feedback and do not pay attention to the content	1	8	2	1	5

Figure 15 counts

Teacher response counts midline and endline (Figure 15)

Have you noticed any positive changes in the content of the feedback you provide to pupils when you use digital feedback since the beginning of the school year?

Midline Treatment Teacher:	**Endline Treatment Teacher:**
No: 4	No: 12
Yes:19	Yes: 7

Research Lead response counts midline and endline (Figure 15)

Have you noticed any positive changes in the content of the feedback to pupils given by teachers using digital feedback, since the beginning of the school year?

Midline Research Lead:	**Endline Research Lead:**
No: 5	No:4
Yes: 11	Yes: 13

Head teacher response counts midline and endline (Figure 15)

Have you noticed, or heard of, any positive changes in the content of the feedback that teachers who use digital feedback give to pupils since the beginning of the school year?

Midline Head Teacher:	**Endline Head Teacher:**
No: 2	No: 3
Yes:6	Yes: 4

Figure 16 counts

Teacher response counts midline and endline (Figure 16)

Do you think the quality of the feedback you provide to pupils has increased over the past school year as a consequence of using digital feedback?

Midline Treatment Teacher:	**Endline Treatment Teacher:**
No: 7	No: 11
Yes:16	Yes: 8

Research Lead response counts midline and endline (Figure 16)

Do you think the quality of the feedback to pupils given by teachers using digital feedback has improved over the past school year as a consequence of using digital feedback?

Midline Research Lead	**Endline Research Lead**
No:6	No: 6
Yes: 10	Yes: 11

Head teacher response counts midline and endline (Figure 16)

Do you think the quality of the feedback that teachers who use digital feedback give to pupils has increased over the past school year as a consequence of using digital feedback?

Midline Head Teacher	**Endline Head Teacher**
No: 4	No: 4
Yes: 4	Yes: 3

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

You may re-use this document/publication (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence v3.0.

To view this licence, visit https://nationalarchives.gov.uk/doc/open-government-licence/version/3 or email: psi@nationalarchives.gsi.gov.uk

Where we have identified any third-party copyright information you will need to obtain permission from the copyright holders concerned. The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Education.

This document is available for download at https://educationendowmentfoundation.org.uk

The Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP

https://educationendowmentfoundation.org.uk

@EducEndowFoundn

Facebook.com/EducEndowFoundn