Education Endowment Foundation

Anglican Schools Partnership: Effective Feedback

Evaluation Report and Executive Summary

February 2014

Independent evaluators:

Durham University

Professor Stephen Gorard

Beng Huat See

Nadia Siddiqui

The Education Endowment Foundation (EEF)

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.

We aim 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;

- Encouraging schools, government, charities, and others to apply evidence and adopt innovations found to be effective.

The EEF was founded in 2011 by lead charity The Sutton Trust, in partnership with Impetus Trust (now part of Impetus – The Private Equity Foundation), with a £125m grant from the Department for Education. With investment and fundraising income, the EEF intends to award as much as £200m by 2026. Together, the EEF and Sutton Trust are the Government-designated What Works Centre for Improving Education Outcomes for School-Aged Children.

For more information please contact:

Robbie Coleman
Research and Communications Manager
Education Endowment Foundation
9th Floor, Millbank Tower
21-24 Millbank
SW1P 4QP

p: 020 7802 1679
e: robbie.coleman@eefoundation.org.uk
w: www.educationendowmentfoundation.org.uk

About the evaluator

The project was independently evaluated by a team from Durham University led by Professor Stephen Gorard.

Stephen Gorard is Professor of Education and Well-being, and Fellow of the Wolfson Research Institute at Durham University. He is a Methods Expert for the US government Institute of Education Science, member of the ESRC Grants Awarding Panel, and Academician of the Academy of Social Sciences.

His work concerns the robust evaluation of education as a lifelong process, focused on issues of equity and effectiveness. He regularly gives advice to governments and other policy-makers, including oral and written advice to the House of Commons Education Committee every year since 2003. He is also a widely read and cited methodologist, involved in international and regional capacity-building activities, and used regularly as an adviser on the design of evaluations by central and local governments, NGOs and charities. He is currently an evaluator for the European Commission Directorate-General for Education and Culture, the Department of Work and Pensions, the Food Standards Agency, the Learning and Skills Information Service, and the Education Endowment Foundation. He is author of nearly 1,000 books and papers.

Contact details:

Professor Stephen Gorard
School of Education
Durham University
Leazes Road
DH1 1TA

p: 0191 334 8419
e: s.a.c.gorard@durham.ac.uk

Contents

Executive summary

The project

This pilot project focused on improving teachers' understanding and use of effective feedback. Participating teachers tried to incorporate feedback into their lessons to help pupils understand their learning goals and become able to develop strategies to reach them. The project employed a cyclical action research design, through which teachers reviewed academic literature on effective feedback before developing ways to apply it in the classroom. The project took place over one school year and involved nine treatment and five comparator schools in the London Borough of Bexley. All pupils in Years 2-6 took part in the study.

Existing international research suggests that improving the quality of feedback in the classroom has the potential to improve learning significantly. However, studies also highlight the difficulties in improving feedback in practice, and there are few clear examples of how to improve feedback in English schools. This project sought to develop a way of improving feedback led by schools.

The pilot evaluation had three aims. First, to assess the feasibility and promise of an approach to improving feedback which required schools to review, understand and apply research findings, including academic papers. Second, to provide formative recommendations that could be used to improve the approach in the future. Third, to provide an initial quantitative assessment of the approach's impact on academic attainment that could be used to inform any future trial.

What did the pilot find?

The approach is feasible and there are some indications of promise. All nine schools completed the action research programme and at the end of the year many staff were receptive and enthusiastic about the approach. A number of good lessons with clear use of feedback strategies were observed. However, in common with existing studies on feedback, there was wide variation in the way that strategies to improve feedback were used.

Many teachers found it difficult to understand the academic research papers which set out the principles of effective feedback and distinguished between different types of feedback. For example, the literature on feedback draws an essential distinction between feedback targeted at the self ('*Great sentence; you are a superstar!*') and feedback which promotes self-regulation and independent learning ('*You have learned some adverbs today. Check if you could add some adverbs to improve your sentences.*'). However, it was not clear in observed lessons that this distinction was consistently understood. Some teachers initially believed that the programme was unnecessary as they already used feedback effectively.

The pilot produced valuable formative information for a potential future project. In order to improve the consistency of the approach employed it is recommended that staff be provided with a large number of examples illustrating the variety of types of feedback. Video recordings of effective lessons could be used as a training resource. This approach would be likely to be more successful than one which required teachers to work from undigested evidence reports. The process evaluation also identified the need for more differentiation in the use of feedback, and a clearer explanation of the use of success criteria in lessons.

The estimated impact findings showed no difference between the intervention schools and the other primary schools in Bexley in terms of annual progress towards Level 4 at Key Stage 2 or in terms of value-added progress scores. However, due to the non-random nature of the comparison and the small number of schools involved it is difficult to draw conclusions with this. The results should not be confused with those of a full trial. Pupils eligible for free school meals made more progress in participating schools than in comparison schools. However, these findings are based on much smaller numbers and so even greater caution is required.

Question	Finding	Comment
Was the approach feasible?	**Yes**	All schools completed the project.
Is there evidence of promise?	**Mixed**	Many good lessons were observed but teachers struggled to understand and use evidence on effective feedback consistently across all schools.
Is the approach ready for a full trial?	**No**	Further development is required to refine the approach and provide more support to make research accessible to teachers.

How was the pilot conducted?

The pilot was a large-scale and in-depth study of 2,000 children receiving the intervention in Years 2 to 6 in nine primary schools. The intervention also took place in the one secondary school in the same partnership but their results are not part of this evaluation. A further 1,000 pupils acted as a partially matched comparator group in five schools. The process evaluation formed the bulk of the fieldwork, with the aim of providing formative evidence on all phases and aspects of the intervention from cascading the training to evaluating the outcomes. Additional data from observation and interviews with staff and researchers, and via focus groups and a brief survey with pupils were collected.

The impact study had a 'before and after' design, measuring the gains made in Key Stage scores using teacher assessment scores. Comparisons were made with results in five other local schools identified by the project lead; therefore the results must not be mistaken for those of a trial. The school-based research approach meant that causal influences could not be robustly identified; the quantitative component of the study primarily sought to provide an estimated effect size for any intervention that could be used in future trials.

How much does it cost?

This is a whole school intervention, involving 10 schools and around 4,000 pupils at a cost of around £88,000. The cost per pupil is approximately £22. This estimate includes the cost of delivering the intervention to nine primaries, and one secondary school not involved in the evaluation.

Key Conclusions
1. Effective feedback has shown promise in previous studies, but this evaluation demonstrates that improving feedback consistently is challenging.
2. The approach appeared to be most effective when training was communal and when objectives and methods were shared. It was least successful when teachers were unclear about the differences between different types of feedback, and when pupils were unable to set clear success criteria.
3. Teachers often struggled to interpret, understand and apply findings from academic research.
4. The study did not seek to assess impact on attainment in a robust way. However, the attainment data which was collected indicated that there may be some evidence of promise for students eligible for free school meals.
5. One future step may be to try and develop the intervention into a more structured programme targeted specifically at low achieving pupils and pupils eligible for free school meals. Greater support, including videos of model lessons could be provided to participating teachers.

Introduction

This is a report on an evaluation of the 'Anglican School Partnership Effective Feedback' programme which was piloted in nine primary schools and one secondary in the London Borough of Bexley over one academic year from September 2012 to July 2013.

Intervention

The intervention being evaluated here was only partly formed at the initial stage. It began with a theoretical understanding of the power of effective feedback based on John Hattie's model of effective feedback, relying on a meta-analysis of meta-analyses (Hattie and Timperley 2007). Using an action research design, the intervention aimed to formulate and develop a fuller programme to encourage teachers to use effective feedback routinely in the classrooms. The programme was shaped in practice as it goes along with each cycle. The intervention was therefore a template for practice at this stage.

The principle behind the intervention was taken from Hattie's model and the idea was adopted and adapted by the 10 schools in the Anglican School Partnership. The project was a whole school intervention involving teachers coming together to decide what effective feedback looked like, and sharing experiences of success and challenges. It consisted of a 'spiral of steps' or action research cycles. At each cycle, the school leads met to plan and reflect on the processes involved. Teachers within schools also met to reflect on their own experiences and share them with school leads.

Hattie's model of feedback is based on the understanding that in order for children to be effective learners they need to bridge the gap between what they already know (prior knowledge) and what they need to learn (desired goal). To achieve this, students need to increase their own effort by first being able to identify their own errors (self-feedback), and use or be taught to use better strategies to complete a task or solve a problem (self-regulation). Teachers can assist in helping students to narrow the gap by giving them challenging and specific goals, clarifying goals and creating the right learning environment. To do this, teachers help students by:

- Identifying the learning goals (or success criteria). These goals need to be specific and challenging;
- Providing information on how pupils are doing and how they can do better. This aspect is called "feed-back";
- Helping pupils to identify further learning possibilities. This aspect is termed "feed-forward".

The model thus proposes three feedback questions:

- Where am I going? This relates to identifying learning goals;
- How am I going? This relates to clear information about performance and the success/failure on a specific task;
- Where to next? This relates to offering pupils information that will lead to further learning.

According to Hattie, the effectiveness of feedback is determined by the levels at which feedback is directed. There are four levels of feedback. These are:

- *Task level*
For example, the teacher may say, 'this is correct' or 'this is incorrect'. It may also include giving directions. For example: 'You need to use more descriptive words.'

- *Process level*

For example, the teacher may say, 'You can make your writing more interesting by using the adjectives and adverbs you have learnt so far.'

- *Self-regulation level*

For example, the teacher may say: 'You have learnt some adverbs to describe how people walk and talk. Check that you have used some of these words in this essay.'

- *Self-level*

This kind of feedback is at the personal level and unrelated to task performance. Examples of such feedback include, 'You are a superstar'.

Hattie argues that feedback at self-level is the least effective in improving performance. The aim is to drive feedback to self-regulation level and to develop self-efficacy and independent learning. However, some instructions are still needed for ineffective learners, for example, for concepts that are difficult to grasp. In other words, the use of feedback is not a substitute for classroom instruction. It is when effective classroom instruction is supported by the effective use of feedback that learning can be enhanced.

Background

Existing evidence for the intervention

This idea for the intervention came from an early version of the Teaching and Learning Toolkit, produced by the Sutton Trust and the Education Endowment Foundation, which at that time suggested that the effective use of feedback had a strong impact on pupil attainment with an indicative effect size of 0.62. The programme follows quite closely Hattie's model of feedback taken largely from his paper: The Power of Effective Feedback (Hattie and Timperley 2007). Although Hattie's model of feedback has not been tested or trialled on a large scale and in real classroom condition, there was widespread belief in its efficacy and it has been promoted to schools in England. This model forms the evidence base for the project. The strength of this evidence based on Hattie's work and those of others on similar topics are discussed here.

From their review and meta-analysis of meta-analyses, Hattie and Timperley developed a model of effective feedback, the purpose of which was to narrow the discrepancy between a pupil's level of understanding and performance and their goal or success criteria. The role of the teacher was to assist pupils in formulating their success criteria, ensuring that they are clear and achievable.

Hattie's model was based on an impressive number of meta-analyses (n=74), involving a large number of studies (n=4,157) and 5,755 effect sizes relating to feedback (Hattie & Timperley 2007). Hattie & Timperley (2007) reported wide variations in effect sizes depending on the types of feedback used, but there was no comment on the quality of these studies and the reliability of the evidence. The study that most informed their model was the one by Kluger and DeNisi (1996) because this study, according to Hattie & Timperley was 'the most systematic', and 'included studies that had at least a control group, measured performance, and included at least 10 participants'. This suggests that studies in the 73 other meta-analyses may not have been so systematic, had no control group, had fewer than 10 participants or did not measure performance. How many of such studies were in these meta-analyses was not known. Therefore, the number of studies whose evidence can be relied on is unknown.

It is also worth noting that many of the studies in Kluger and DeNisi's meta-analysis were not classroom-based. Presumably these studies were undertaken in a controlled or laboratory condition. How would the results compare in real life classroom situations which normally have about 30 students and other atmospheric distractions, such as noise from outside the classroom, and interruptions from other students? Classes with SEN children can also affect classroom delivery

because of the attention needed. The way an innovation is implemented in a real classroom situation (fidelity to treatment), the experience and expertise of the teachers, and the age of the children can all contribute to the success or otherwise of an intervention (reliability and external validity). None of these were discussed.

The paper went on to say that its evidence was based on 131 studies (conducted largely in control conditions) and included 470 effect sizes. Hattie calculated the average effect size from these studies as 0.38 (SE=0.09). Of these, 32% showed negative effects. This means that 150 of the effect sizes were negative. Moreover, these were also small-scale studies involving on average fewer than 100 participants – or under 50 in each arm. Hattie and Timperley reported that the 'average sample size per effect was 39 participants'. Neither was it clear how the impact of such an intervention differed for children from disadvantaged backgrounds and for children of different ages. Clearly, very young children could have difficulty in setting their own goals. Would this intervention be appropriate for five- and six-year-olds?

The evidence of impact in Hattie's model can thus be said to be unclear. It is based on a summary of passive rather than active research designs. The meta-analyses used different calculations of effect sizes, for different measures of the same parameters (e.g. different types of reinforcement and a range of feedbacks) for different groups of children of different phases of schooling. Some studies were specifically for SEN children, or children with behavioural, emotional and disruptive behaviour. How the authors arrived at the effect sizes that they did in the paper was not explained. They assume that there is a standard gauge for effect sizes, which there clearly should not be (Gorard 2006). Neither do they link these effect size benefits to their costs, nor link them to unintended and disadvantageous consequences. Looking at the studies that Hattie cited in his paper, it is not always possible to locate the effect sizes listed in the summary table (Table1, p. 83). In summary, it is not yet clear if Hattie's model of effective feedback works in real classroom conditions.

Another well-known study on the use of feedback is by Black and Wiliam (1998), called "Inside the Blackbox". They reviewed studies on the effects of formative assessment which they defined as any activity by teachers and pupils that provided feedback to inform teaching and learning. Their summary, which was built on an earlier review of 23 studies by Fuchs and Fuchs (1986), included 20 further studies. All the studies reviewed indicated a substantial impact of formative assessment for the learning of pupils of all age groups (age five to undergraduates) and across subjects and nationalities. Black and Wiliam found that the average effect sizes of the impact of formative assessment experiments on pupils' attainment ranged between 0.4 and 0.7. While the Fuchs and Fuchs study found that formative assessment was particularly effective for children with special educational needs, Black and Wiliam found that formative assessment was effective for low achievers more than other students. Others have suggested that this approach will not always be effective, perhaps especially if rolled out without due care (Smith and Gorard 2005).

At the heart of Black and Wiliam's programme is the role of the teacher. Teachers need to know how their pupils are doing, and the difficulties they face, in order to tailor their teaching to meet their pupils' needs. This knowledge is then used to modify teaching. Successful FA builds on pupils' self-esteem, focusing on specific problems with their work, with clear explanation on where they have gone wrong and how to correct it. One important feature of successful FA is the ability of pupils to set goal-oriented criteria, or what Hattie termed, success criteria. Pupils assess their own progress, identify areas that need improvement and understand strategies required to achieve this. However, Black and Wiliam stressed that for this to work, students need to be trained to assess themselves and understand what they needed to learn. Classroom and homework tasks were structured to include opportunities for pupils to communicate their understanding of their learning objectives. This could be through discussions, observations of activities or though written work. In this way teachers receive feedback about their pupils' learning process.

Further small-scale evidence for the impact of using success criteria comes from the work of White and Frederiksen (1998). Their study examined the use of reflective assessment in the teaching and learning of physics for children from Grade 7 to Grade 9. The intervention involved telling students explicitly the criteria that would be used for judging their work. Students then used these criteria to

evaluate their own work and that of their peers. The quality of their projects was then evaluated by teachers. These criteria are similar, in some senses, to what Hattie calls 'success criteria' or what schools call 'Learning Outcomes'. The authors report that students who were able to assess their work produced higher quality work as judged by the teachers than those who did not. The effect was greater for low-achieving students (effect size of 1.0) than for high-achieving students (effect size of 0.27). However, the study was based on only three teachers in two schools. The report does not explain how many classes were in each treatment group (only about 60 students appear in the key analyses). The group were matched not randomised, yet the authors conduct their analyses using ANOVA. Despite a range of pre to post differences, a subsequent test of physics knowledge showed that students in all the classes and in all grades performed the same, suggesting no effect from the intervention.

It is also worth noting that there is a difference between the Reflective Assessment process and the type of feedback advocated by Hattie and Timperley. In Reflective Assessment the success criteria is made known to the pupils, whereas in Hattie's model, students set their own success criteria. There is therefore a risk where students set criteria that are too low or which are inappropriate. A lot of guidance would be needed from the teacher.

Explanation of the stage of development of the intervention

Thus the existing literature has suggested that enhanced feedback is effective, but the results are sometimes based on studies with flaws, or an inappropriate synthesis of designs involving different phases and measures. The Anglican School Partnership project is the first UK pilot trial to evaluate Hattie's model in real classroom conditions across a range of subjects, age groups, and in a number of schools. It is, to a certain extent, also a feasibility trial to see if such intervention could be carried out by schools themselves with teachers coming together to field test an idea previously tested under controlled conditions in many studies or tested in only small-scale studies, and to use the lessons learnt from this pilot to work towards a full test of effectiveness in future years. The evaluation is for the pilot only, but takes into account its formative nature, adopting a design experiment approach. The pilot is one phase of a larger design study working towards a trial, but retaining the flexibility of a pragmatic template intervention that practitioners can adapt to suit their context and needs. The evaluation should yield results that will help the design of a future trial, and also provide guidance to EEF and others on the viability of the action research and design experimentation models.

Details of any relevant policy or practice context

The concept of effective feedback, and the setting of success criteria or learning outcomes, is not new. The policy background emphasising the use of formative assessment in the UK began with the introduction of the National Curriculum (NC) into primary schools in 1989, which emphasised the formative aspects of assessment and the use of a range of assessments (House of Commons 2009). One of the aims of the NC was to make expectations for learning and attainment explicit to pupils. This is the principle of goal-oriented criteria advocated by Black and Wiliam in their Black Box experiment and in White and Frederiksen's reflective assessment study. In this respect it is also similar to Hattie's success criteria.

Objectives

The intervention evaluated here is a pilot trial of Hattie's model of effective feedback that has been evaluated in a number of studies but largely in controlled conditions (Hattie and Timperley (2007). Similar models have also been trialled in small scale experiments, such as in Black and Wiliam (1998)

and in White and Frederiksen (1998). The question is whether the programme is practical for teachers to use in the classroom, and whether it is effective in changing teacher behaviour. The project is, therefore, not to develop an 'effective feedback intervention'. It does not specify what teachers do. Instead teachers are meant to try using the research findings in their classes, learning from this, and reporting back. The aim is to come up with a programme through which groups of teachers and teaching assistants are encouraged to review the evidence about feedback and work together to develop concrete examples of how they can apply this in the classroom.

The impact evaluation is to estimate the impact of the proposed intervention on pupils' attainment, with a special focus on FSM-eligible and other disadvantaged pupils. There is no true counterfactual and comparison is made with results in other local schools identified by the developer, and with the published results of all other primary schools in the same local authority. As such, the impact evaluation is largely to provide an estimated effect size for any intervention that could be used in future scaled-up trials. The aim of the process evaluation is to provide formative evidence on all phases and aspects of the template intervention from cascading the training to evaluating the outcomes. It will involve the perceptions of participants including any resentment or resistance, and lead to advice on improvements and issues for subsequent scaling up.

Schools themselves also evaluated the progress of their pupils at the end of each cycle. This progress relates to pupils' understanding of the process of feedback (i.e. setting their own success criteria, assessing their own work, understanding what they need to do to improve and how to achieve this). This is done with the use of a teacher-developed Pupil Learner Effectiveness survey at the end of each cycle. Results from the survey were made available to evaluators who assisted with the analysis of the responses and provided appropriate suggestions and recommendations regarding the design and formulation for any future questionnaire, if the intervention is to be continued or introduced on a wider scale.

The purpose of the process evaluation is thus to assist in improving the template for a later trial, and in deciding whether an action research approach is useful in such circumstances. Therefore, the process evaluation involves gathering additional data from observations and interviews with teachers, school leads and pupils.

Project team

The programme was developed and conducted by the Anglican School Partnership led by the executive head teacher of Trinitas Academy Trust. The school leads from the other nine schools in the Partnership supported the project.

Methods

Evaluation Design

This evaluation was based on action research. A model of feedback derived from Hattie and Timperley (2007) was used to develop the intervention. Teachers' and pupils' understanding of the intervention was constantly monitored and adjusted at the end of each action research cycle (ARC). There were four cycles in all, but ARC 2 and 3 were merged as it was felt that the time interval between the two cycles was too short for any observable progress.

Some changes were made once the project was underway. First, although this was a whole school intervention, it was decided after the first training meeting not to include the reception year and Year 1 pupils in the analysis of outcomes for three reasons. One reason was that school leads and teachers thought that the younger children were not able to set their own success criteria. Many of the younger children also did not quite understand the questions in the Pupil Learner Effectiveness (PLE) survey. Another reason was that many schools did not provide their end of year teacher assessment for Year 1 and reception year. A decision was also made to exclude the only secondary school in the partnership since there was no direct comparison. This was therefore a project for primary age Years 2 to 6. After much effort by the programme developer only five local comparison schools were found who agreed to participate by providing pre and post data.

There were variations in the subject areas targeted. After the first initial meeting of school leads, some schools decided to implement the intervention only in certain subjects, and many chose to employ the strategy in literacy and numeracy lessons.

Eligibility

Participating schools were all those in the Anglican School Partnership in the Bexley area who agreed to take part in the trial. All schools signed a letter of agreement stating that they were happy to take part in the project and in the evaluation. As it is a whole school project, all pupils in the schools were involved. Comparison schools were volunteers from the Bexley area but not in the partnership.

Intervention

The project was a one-year pilot trial, employing an action research design approach, to help develop a programme to enable teachers to engage with the evidence on effective feedback and to incorporate this routinely in their classroom teaching. The intervention involved teachers engaging with pupils via the use of feedback strategies to help them to understand their learning goals and to use these to develop strategies for their learning.

The project began with initial training for school leads, heads of schools and head teachers, based on reading and discussion of the paper by Hattie and Timperley (2007). Subsequent to this, a moderation meeting with school was convened. School leads then delivered feedback moderation training to staff, followed by a moderation staff meeting to establish a starting point. Schools collected real examples of feedback at the four levels in Hattie's model. A training pack with training materials was prepared. Working in pairs, schools received training on an INSET day. Learning teams established starting points by carrying out audits including pupils' baseline data. The Pupil Learner Effectiveness survey

and a feedback survey using a feedback grid were administered. All pupils completed the teacher-developed online survey of Pupil Learner Effectiveness aimed at identifying pupils' starting points and their learning strategies. This was the starting point.

The project involved four cycles of action research, named Action Research Cycle 1 (ARC 1) to ARC 4. At the end of each cycle there would be a School Lead Evaluation meeting to share examples of good practice.

ARC 1

Each school lead was given three days of supply cover to collect examples of Hattie's three types of feedback (feed-up, feedback, feed-forward) and the four levels for each type (personal, task, process and self-regulating). Teachers also identified 'Where pupils are going', 'How they are doing' and 'Where to next'. Teachers audited each other's' lessons to look for prevalence of the 12 combinations of feedback. They then created an action plan to try out a new balance of feedback, making it 'proportionate' to its value, aiming for more self-regulatory feedback and fewer personal comments.

This was followed by three further cycles of action research to move pupils from perhaps having the characteristics of ineffective learners (e.g. not planning) to effective (e.g. planning) learners. Essentially an effective learner is one who knows where they are going and how they are doing, and what they need to do to reach their goals. So students start by setting their success criteria. They need to be able to identify their own mistakes and know what to do to correct them. This is the self-regulatory stage.

ARC 2 and ARC 3

In Cycles 2 and 3 each teacher audited pupils' skills in terms of Hattie's model of being an effective learner using the results of the survey as a starting point. Teachers reflected on their own practice and monitored pupils' understanding and application of the concept of an effective learner. Schools identified areas for improvement and suggested strategies to achieve these. School leads met to discuss issues and challenges and reported on progress made. At the end of ARC 3 a second Pupil Learner effectiveness survey was conducted.

ARC 4

ARC 4 was quite a short cycle given the end-of-year assessments and other end-of-year activities. Results from the second survey were analysed and compared with those in the first survey to assess progress. Teachers met to discuss issues highlighting successes and challenges. Teachers also observed each other's lessons, collected and shared examples of feedback used and strategies employed to move pupils towards being a more effective learners. This was followed by a school leads evaluation meeting where schools reflected on their experiences and shared examples of good practices.

None of these ARC activities nor the training took place in the five local comparator schools, which merely provided the pre and post data on their pupils.

Process evaluation methods

The process evaluation was conducted by the independent evaluators in collaboration with the programme developer. The latter included the overall project lead supported by school leads from each school. The project lead conducted the training of school leads and they in turn trained the

teachers. Together they monitored the intervention, held review meetings and revised the procedure along the way. They also coordinated the collection of formal records and feedback from teachers and pupils. The evaluators attended training sessions and review meetings to observe the delivery of the training, and assess the training materials, noting staff reaction to training and fidelity of training in cascade. Evaluators also made school visits to observe the implementation of the feedback strategies in the classroom. On these visits informal interviews with teachers, students and school leads were arranged. These interviews were conducted without a formal structured schedule, although a general guideline with sample questions for interviewing teachers was provided (see Appendix A). The observations of staff training and implementation of the programme in action were as simple, integrated and non-intrusive as possible. The schedule of visits was agreed with the school leads and interviews were arranged at that time. Classes and teachers identified for observation and interviews were selected to represent a range of year groups covering KS1 and KS2. The project developer also suggested further review meetings and additional training sessions that evaluators could attend. In total the evaluators made about 24 trips to the research sites.

The aim of the evaluation was to provide formative evidence on all phases and aspects of the template intervention from cascading the training to evaluating the outcomes. The purpose of this was to improve the template for a future trial and to test the feasibility of the action research approach for such a trial. Thus, a substantial part of the evaluation fieldwork was to assess how closely schools adhered to the intended intervention, and what the short term or intermediate impacts were (such as changes in classroom interaction).

The basic idea of this action research was that actions (interventions) were evaluated formatively in context, constantly monitoring and revising the procedures while live. This meant checking for changes in consequences (effects of the action) over and above what might otherwise have been expected, learning what seemed to work best and what the barriers were, modifying the action for the next step in the cycle, and starting again. For this reason, the process evaluation looked specifically for data that addressed:

- the reaction to training
- the fidelity of training in cascade
- whether the teams understood the process and purpose
- the contents and use of the starter pack
- starting point and subsequent assessments
- how missing data was handled
- changes in classroom interaction
- how pupils took control of their own 'feedback' loop in improving evidence-informed practice
- audits by classes of teacher feedback, and learner effectiveness
- the ongoing 'engineering' of a pack and web resources
- whether teachers could tell if the template is working, or modify it accordingly
- whether there appears to be an impact on how children are learning
- whether teachers provided useful and better feedback
- and whether pupils responded to feedback.

Impact evaluation

The impact design was a before and after study with a convenience sample of nine schools (only the primary schools were involved in the evaluation) and a partly matched comparator group of five local primary schools. The longitudinal approach followed entire cohorts through one year of schooling, intervening, monitoring and adjusting the intervention as the programme progressed. The action research approach, however, is not ideal in terms of identifying causal influences as there is no true counterfactual, so outcomes (pupils' performance at end of year teacher assessment or at KS1 and

KS2 assessments) were compared to the progress of cohorts in comparison schools in the Bexley area but not in the intervention partnership. There was also a comparison between the results and progress of disadvantaged pupils (FSM or eligible for pupil premium) and the rest. However, none of these approaches sought to provide a scientifically defensible comparison group to calculate (rather than estimate order of magnitude of) the 'effect size' of the intervention. This was acceptable for the pilot, but any follow up must involve randomisation to treatment or a control.

Outcomes

The primary outcome measure for all pupils was the fine points scores for teacher assessment or Test in the appropriate Key Stage for all relevant year cohorts (such as Year 6), and progress from the equivalent scores from the previous year. Sub-group analyses of pupils by FSM-eligibility, sex, EAL and SEN were also carried out to assess the impact for potentially disadvantaged children as defined by these measures. None of these measures or variables was additional to those collected routinely or as part of the proposed intervention. They are appropriate and standardised (as far as is possible).

Pupils' prior background and contextual data were provided by the SIMS records held in each school. These included Key Stage results (levels and points), sex, month of birth, FSM status, SEN status, ethnicity, and first language. Supplementary pupil data, such as individual attendance records, date of leaving (if during the project), and any disciplinary records such as suspensions or exclusions (where applicable) were also collected. These came from existing school records.

In addition the value-added scores and Year 6 Key Stage 2 results for 2012 and 2013, for all schools in Bexley, were obtained from the DfE Performance Tables website. These provided a larger and more robust (in terms of assessment) comparison for Year 6 pupils.

Sampling and recruitment

The treatment schools were those forming the Anglican Schools Partnership in Bexley. They had all agreed to take part in the study as full partners. These included nine primary and one secondary school. It was intended that a further 10 comparison schools from the same local authority would be recruited and matched on available measures of school organisation and intake. These would be used to provide context and pre- and post-test data as a comparator group not receiving the intervention. Together this would provide data to estimate the likely effect size for a trial. However, in practice, recruitment of comparison schools was difficult, and after much effort by the lead developer, only five other schools in the Bexley area which were not in the partnership agreed to take part in the study. These were all primary schools. As there was only one secondary school, this meant that there would not be a direct matched comparison school. For Year 6 only, the results of the nine treatment schools are also compared with the published results of the other 59 primary schools in the same local authority.

Not all year groups contributed to the estimated effect size (see below). Therefore, in this evaluation the focus is on the nine primary schools, and on KS1 and KS2 students only. Data for reception class students are not included in the evaluation. This is partly because prior attainment measures are not available for this age group of children, and also the use of self-regulatory feedback at process and task level and the setting of success criteria may not be relevant to them.

Letters of agreement were sent out by the programme developer to participating schools. Schools agreed to the evaluation when agreeing to participate in the intervention.

<u>Allocation to groups</u>

Allocation was pre-determined as all the schools in the Anglican Schools Partnership would be participating. This was how the intervention was designed when it was funded initially.

<u>Analysis of outcome measures</u>

An attempt was made to get complete test scores for all pupils even where they were initially absent or left the schools during the study. Such results are analysed in terms of their original schools (intention to treat). Differences are calculated for the gain scores from the prior Key Stage scores to the subsequent Key Stage scores, and presented in standardised form as Hedges' g 'effect' sizes. The results are presented overall, by years and subjects (reading, writing, maths). Differences that appear under all of these conditions would be considered robust or substantial. Sub-group analyses include boys and girls separately, and for FSM-eligible pupils only.

The published Year 6 KS2 results (percentage attaining Level 4 or higher in reading, writing and maths, and the VA KS1 to KS2 progress scores) for all schools in Bexley were divided into those of the treatment schools and all others. The results were averaged for the two groups, weighted by the number of Year 6 pupils in each school. There were 422 Year 6 pupils in the treatment group and 2,187 in the other primary schools in Bexley.

Three regression models were also created using combined year groups, one with each of the subjects (English, maths, science) as the 'predicted' variable. Potential predictors were entered in two blocks. The first block included the prior attainment for the same subject, and individual pupil characteristics such as FSM, sex, ethnicity, SEN and EAL. The second block consisted of knowledge of the treatment or comparison group. In this way, the model can suggest the extent to which the treatment could have an impact once other known factors are accounted for. However, it must be recalled that this is primarily a formative evaluation, and that these calculations are to provide an estimated effect size for any future trial, and to rehearse and pilot the data requirements. They do not have the authority of a trial.

Process evaluation results

Training in cascade and review meetings

The training was generally well conducted, although the content may have been unclear to some teacher participants (see below). It progressed through the four cycles, starting with the training of school leads, followed by the training of teaching staff. Additional training sessions were also conducted by the programme developer when it was noticed that many teachers were not adhering to the programme as suggested. The aim of these sessions was to inform and remind teachers about Hattie's Effective Feedback Model and the strategies to achieve impact.

To ensure that the staff across all the schools received similar training, the programme developer and school leads had a prior agreement on what procedures and practices would be adopted. The training for school leads and subsequent training for teaching staff followed the same format. This included the use of 30 PowerPoint slides, reading of Hattie and Timperley's article followed by discussions and presentations from the participants.

There was some resistance to the programme at the initial training session. Some school leads thought that the feedback strategy was something that they had already been using in the class anyway. Some also disagreed with Hattie about the likely negative effect of praise at self-level:

'We do this already in our classroom.'

'We get the grades so why do we need to do this?'

'It's just another intervention isn't it?'

After the initial training, teachers appeared more receptive of the programme and agreed with the potential benefits of the intervention:

'We often have lots of initiatives but it's great to focus on feedback.'

'We are learning about how to improve feedback which will support students in their learning.'

'I think it's really good, it matches with what we did at university so it definitely helps to recap it.'

A number of teachers suggested that the inclusion of examples of the different types of feedback and modelling of feedback styles would have improved the training.

Review meetings

Review meetings were held at the beginning and end of each cycle to reflect, share and moderate the strategies. There were also regular review meetings of programme developer and school leads to share experiences, barriers and successes. School leads also met regularly with staff to give them feedback as well as receive feedback from them regarding what needs to be done to improve.

Subsequent review meetings appeared to be more about the strategies and how schools implemented these strategies. These however, were less about the use of feedback, but more about how teachers got pupils to write and set their own success criteria, such as using colour coding, symbols,

emoticons, getting pupils to write success criteria in their notebooks or have them displayed in the classroom. This was a bit problematic as the concept of success criteria was not accurately understood by most pupils. There was little discussion of examples of the different types and levels of feedback to use and how they could be effectively used in the class. This was perhaps the weakest part of the preparation.

Contents and use of starter pack

The starter pack included the 30 PowerPoint slides handouts, the article by John Hattie and a chart of the feedback strategies. Teachers found the article heavy and academic. Common comments included:

> *'I need a translator to understand what this article is saying.*
> *I just cannot understand what he [Hattie] means and what he wants us to do.'*

> *'I don't understand what we are meant to be doing.'*

> *'What do they mean by 'process'?'*

The article was a meta-analysis of studies where the authors established a strong case in favour of feedback and developed the categories of feedback that had positive impact on the learning process. It was an academic paper and there were few, if any, actual examples of each type of feedback strategies advocated by the authors, which teachers could adopt and implement in the classroom. The paper was too dense for practitioners to read and use in the time available.

The starting point and the Pupil Learner Effectiveness Survey (PLE)

The Pupil Learner Effectiveness Survey was used as a starting point to establish pupils' learning strategies. The PLE survey was an online self-assessment questionnaire of pupils' learning process using Survey Monkey, created by the developers and further developed based on input from teachers and head teachers (a version appears in Appendix A). The survey was administered at the beginning of the cycle as a starting point and again at the end of the cycle. Results were compared to see if there were any changes in pupils' approach to learning.

Concerns were raised by teachers about the questionnaire. Among these concerns was the language of the survey. The phrasing of some question items was vague and ambiguous. Teachers were also concerned that the very young children (those in the early years) and those whose first language was not English might have difficulties with terms like, 'excellent learner' and 'success criteria'.

Teachers' understanding of the process and purpose

Schools were clear about the steps involved, such as using the PLE Survey responses to identify priorities and targets. They understood that the aim was to move pupils towards self-regulation to become independent learners. However, it was not clear if all teachers and school leads similarly understood the feedback strategies advocated by Hattie. It was also not clear if all teachers understood the difference between higher levels of feedback and the lower, less effective and even negative types of feedback. Some teachers thought that they were already using the types of feedback suggested, and that it was what any good teachers would already be doing anyway in their classroom. One teacher also said they did not understand what 'process' feedback was. In the first

training session some teachers clearly indicated that they did not know what the different types of feedback were. By ARC 4 it was still not clear if they had fully understood.

Teachers' use of feedback strategies in the classroom

There were some excellent innovative and creative lessons where teachers modelled to pupils how to correct their mistakes. This was quite an effective strategy to move towards self-regulation. However, many of the lessons observed were much poorer. There were lessons where teachers simply got pupils to self-evaluate or peer evaluate. The successful lesson usually included teacher monitoring, checking that pupils understood and were doing the right thing by actually going round to look at pupils' answers or asking pupils to say what was right or wrong about the answers (feeding back to teachers), rather than just asking pupils to tell the teacher if they had understood. For example, in some lessons, although pupils were asked to fill in their success criteria and how these could be achieved, some pupils had not filled them in and on some tables none had completed the task beforehand. At the end of the lesson all pupils claimed that they had met their success criteria even those who patently had not, or who did not even have any SC. There was no reciprocal feedback from the pupils to the teacher as the teacher was not aware that pupils had not completed their forms. Perhaps pupils did not know how to. This was not checked. According to Hattie, feedback from pupils to teachers is at least as powerful as teacher's feedback to pupils. It signals to teachers what pupils have learnt or have not learnt, and what they need to do next.

Lesson observations by evaluators did not pick up clear evidence of the use of the different levels of feedback – either orally or written. What was clear was that teachers were using success criteria and lesson objectives. And even then it was not clear if teachers and pupils alike understood what SC meant. Much of the feedback used in the classroom was still at task level and self-level, and occasionally process level. There was hardly any feed-up or feed-forward. There was still a lot of use of praise. In one observation throughout the lesson the teacher was making comments such as:

'You're very good at turn-taking.'

'Well done children, you are working well.'

'Well done, you're meeting one of your targets by reading aloud to the poem.'

'You're a superstar, yes it has repetition in the poem.'

Even though such feedback had been identified as ineffective and in some instances even harmful, by the research used as a basis for the intervention, teachers were still using it a lot in the classroom. According to Hattie, praise has a place but to praise when achievement was not really warranted (e.g. reading aloud to a poem or stating that there was repetition in the poem) may not be encouraging the right kind of learning behaviour. In many instances, even when pupils were praised, it was not made clear to pupils what they were praised for. Praise was not directed to learning. It was couched in vague terms like, 'lovely', 'good' and 'great'. Because teachers were not specific in their feedback pupils did not know what was good and what they need to do to improve.

There were attempts to move pupils to self-regulatory feedback, but it seemed that some teachers were struggling to do this. Comments like: 'Check your answers again', 'Look at your work again', 'Work out the answer yourself' were obvious attempts at self-regulation, but they were sometimes not specific enough to guide pupils to self-regulate. For example, pupils may not know what was wrong with their answer or what to look for. More guidance may be needed. One teacher told pupils that: 'To be successful we will need to talk to our partners, work sensibly and share ideas'. There was generally still a lot of teacher talk. School leads felt that there should be greater pupil participation and less teacher talk. Teachers felt that they needed more time to consolidate feedback styles and teaching approaches.

School leads reported that the inability of some teachers to differentiate between the effective and less effective learners had resulted in the more able pupils being 'ceilinged' – not challenged and stretched. In a number of lessons observed, the less able pupils seemed to get more attention from the teacher assistants and teachers, while the more able were not supported. Setting meant that there was less of a chance for the less able pupils to model the effective learning strategy of the more able pupils. Some school leads proposed that mixed ability classes might be a better option.

In summary, the evidence was not very clear whether teachers in general were able to take control of their own 'feedback' loop. There were some good lessons with clear use of feedback strategies linking pupils' success criteria. It was difficult to say if these were effective teachers to begin with or whether the intervention had made a difference. The less successful lessons tended to be those where teachers made fundamental pedagogic mistakes, like not checking pupils' understanding or not personally checking/monitoring pupils' work. This could be more to do with the individual teacher's instructional competence. It was also not clear if teachers understood the different levels and processes of feedback. Part of the reason for this could be the lack of examples and modelling for teachers to practice.

Impact on pupils' learning

As the project progressed, there was greater acceptance and more enthusiasm, with some teachers reporting seeing effects.

Pupils' response to the programme

School leads reported that children were excited and enthusiastic especially with setting their own success criteria. Children were expected to work out the answers for themselves and only sought help when absolutely necessary.

Developing resources

One of the aims of the project was to develop resources for any future trial. This included a website for sharing resources and examples, with a forum for teachers to share their experiences. Because the project had already started with the use of Dropbox, the idea of the website was abandoned for part of the project. However, some school leads thought the website could also provide a forum for pupils' voices to be heard. With the cooperation of the school leads, the evaluators developed the website, with schools providing the materials to be put on the site. As the project was unable to appoint an administrator to manage the site in the short timeframe available, the website did not develop as intended. If the trial was to be scaled up, the potential of using the website forum for sharing experiences and resources could be explored.

Conclusion of process evaluation

The intervention did not progress as planned. As far as it was possible to tell, the intervention was not widely and readily accepted at the outset, either because the paper that introduced it was not the right way to do it, or because teachers already thought they used feedback. Over time and after training and experience, most staff became more receptive and enthusiastic.

Barriers to effective delivery of intervention

Lack of feedback examples

There were not enough examples of each type and level of feedback and modelling of use of feedback strategies. There was no common definition or examples of feedback for teachers. School leads reported a lack of consistency among schools in the way the feedback strategies were implemented. It was clear that schools were adopting different approaches. In the training much of the discussion was about strategies implemented, but not really examples of feedback. There was also a sense that teachers did not fully understand the different types and levels of feedback. They may have been using 'feedback' as they understood it, but not the types espoused by Hattie and Timperley. Instead, teachers were making their own understanding and interpretation.

Lack of consistency in interpreting and applying Success Criteria

There were different understandings of what constitutes Success Criteria (SC) and how SCs were applied. By ARC 4 it was still not clear whether staff fully understood what SC meant and what they should look like. In a number of cases, it was apparent that teachers were not clear what SC meant, even by ARC 4. In some schools, SC was a common list developed in agreement with the class, while in other schools individual pupils set their own SC. In one school the SC was explained to the very young pupils as 'Remember-to do'. In others it was interpreted as the learning objectives. From observation it is apparent that where pupils set their own SC, some were not SC and had nothing to do with learning objectives, e.g. 'Record answers in a table' and 'better handwriting'. SC also tended to be very low level targets and sometimes not related to the learning objectives.

In some cases the SC were couched in very vague terms. Although schools reported that their pupils were able to write their own success criteria and to tell when they were able to take it forward, it was apparent from school observation that this was sometimes based on simply asking pupils through a show of hands if they had met their success criteria.

Reliability and validity of the Pupil Learner Effectiveness Survey

The Pupil Learner Effectiveness (PLE) survey was a focal point of the intervention as schools were expected to use the findings as starting points and to inform their decisions regarding the next step. Relying on the findings of the survey for what needed to be done meant that the questionnaire had to be reliable and valid.

Some teachers remarked that they had a feeling that pupils chose the answers which they thought would please their teacher. Teachers noticed that pupils' responses did not often relate to their ability and behaviour. So although pupils may think that there were excellent learners or that they knew how to correct their own mistakes, their teachers did not think they were or could. School leads also reported instances where teachers helped pupils answer the survey questions rather than letting them pick the answer that was most appropriate for them.

Also it was clear from teachers that pupils did not fully understand the questions. The language used was not easily accessible to very young children and to EAL children. These were later exempted from the survey. Also there was ambiguity with terms used. There are doubts as to whether pupils fully understood the phrase, 'excellent learner'. Many pupils took it to mean 'high attainer'. For example, they said 'I am an excellent learner because I can read more words now'. Some pupils thought a tidy classroom was 'Excellent Learning' as they said if the classroom was tidy then they would know where to look for things. Analysis from the survey results showed that the younger pupils (those in Year 1 and in Infant schools) were more likely to report themselves as 'Excellent learners'. This could be a

positive thing as it suggests pupils' awareness of their own limitations. Younger pupils, not being exposed to tests and national assessments yet, do not routinely have their work judged, so may have been less aware of their own weaknesses.

Outcomes

A website specifically for the Anglican School Partnership was initially developed for schools to post and share good examples of effective feedback. There was also a forum for students to post their comments. The website was not maintained following the completion of the project.

No unintended outcomes were noted. Some teachers reported that lessons were running better and that their pupils had become better learners.

Fidelity

As a template intervention, precise fidelity is hard to judge. All schools participated as agreed, with none dropping out. The intervention was visible on the walls and in the organisation of each school visited, and evident in each lesson visited.

Formative findings

Recommendations

Although the use of feedback may seem like the most natural thing to do for teachers as they employ it routinely in their lessons, to use it effectively may require skill and practice. One of the aims of this project was to encourage teachers to consciously use higher levels of feedback to encourage learning. Effective feedback must be accompanied by effective instruction. Teachers need to be clear about what success criteria are and what the different processes and levels of feedback look like. Based on the process evaluation, the following suggestions emerge:

- Make available ample examples of success criteria, and different types of feedback. Video recordings of effective lessons could be used as a training resource so that teachers can model these lessons.
- Increase use of higher levels of feedback (e.g. more process feedback).
- Minimise the use of self-feedback, which is least effective. E.g. 'You are a superstar', 'This is a clever idea'.
- Feedback should be clear, simple, specific and directed.
- There needs to be a consistent definition of what success criteria are. Success Criteria should be clear and specific and related to the learning objectives. Success Criteria need to be phrased in specific or measurable terms so that pupils know when they have achieved them or not. They should focus on what students should know and realistically be able to do by the end of the lesson or activity.
- Teachers need to make appropriate judgements about when, how and what level of feedback was suitable for pupils.
- Greater differentiation in the use of feedback. More use of feed-up and feed-forward for the more able pupils to provide the challenge. Hattie and Timperley (2007) suggested that for less able pupils it is more effective for the teacher to provide elaborate instructions than feedback on concepts which are difficult to grasp. Feedback needs to be clearly directed. To quote Hattie and Timperley:

"To be effective, feedback needs to be clear, purposeful, meaningful, and compatible with students' prior knowledge and to provide logical connections."

- Effective classroom instruction must be used in concert with feedback. Feedback should not be a substitute for classroom instruction. For example, telling pupils that they need to use more interesting vocabulary is not helpful if pupils have not learnt the vocabulary. Or saying to pupils, 'check your answers again', if pupils cannot see what is wrong with their answers. Hattie and Timperley noted that in some instances good classroom instruction can be more effective than feedback. Feedback has to be built on something. If there is no initial learning or surface information, feedback is of little use.
- If the programme is to be introduced to other schools or scaled up, it is crucial that teachers are properly trained to use effective feedback and linking success criteria to learning objectives, rather than just making up their own interpretation which may not be accurate.

Comparison group activity

The comparator schools only provided pre and post data for their pupils. They did not adopt the intervention.

Impact evaluation results

Participants

There were nine treatment schools and five comparators. There was no school drop out. All schools were coeducational and in urban areas. Treatment schools ranged in size from 207 to over 670, with between 3 and 21% FSM-eligible pupils. Comparison schools ranged in size from 320 to over 620, with between 7 and 18% FSM-eligible pupils (as shown in the boxes below).

Treatment schools

Type of school	Age range	No. pupils	% boys	% SEN	% EAL	% FSM	% non-white UK	KS2 point scores 2012
VA*	3 to 11	401	47.9	6.2	8.4	8.1	31.6	28.9
Academy sponsor	4 to 11	326	50.6	11.7	31.6	13.5		27
Community	3 to 11	433	53.3	6.5	14.2	17.2	32.6	28.4
VA	3 to 11	438	48.4	4.7	11.7	5.2	30.2	29.1
VC**	3 to 11	677	51.8	3	1.1	2.9	12.8	29.9
Community	3 to 7	235	51.1	8.1	23.4	34	38	
Academy sponsor	4 to 11	233	49.8	14.6	59.1	20.9	72.1	27.4
VA	4 to 11	207	45.4	5.8	2.3	7.2	18.1	31.1
VA	5 to 11	212	55.7	3.3	6	6.1	29.4	30.8

*VA – Voluntary aided
**VC – Voluntary controlled

Comparison schools

Type of school	Age range	No. pupils	% boys	% SEN	% EAL	% FSM	% non-white UK	KS2 point scores 2012
Community	3 to 7	464	50.4	4.5	5.8	11.9	36.8	
Community	5 to 11	420	48.8	5	10.3	8.3		29.8
Academy converter	3 to 11	323	47.1	8.4	24	28	44.6	
Academy converter	3 to 11	625	52.5	3.8	2.7	7.3	12.8	28.4

Type of school	Age range	No. pupils	% boys	% SEN	% EAL	% FSM	% non-white UK	KS2 point scores 2012
Community	3 to 11	475	51.8	11.2	10.6	18.4	38.9	28.8

The number of pupils with achieved data in each year group varied partly because of the age ranges of schools, and partly because some data was missing on individual pupils. In Year 2, one school sent no readable data and a further 61 individual pupils across all schools had some missing pre or post data. This was the year with the most data missing, and it tended to decline with the age group. In Year 6, a different school sent no readable data and a further 20 pupils had some missing pre or post data. Most of the individual missing data by Year 6 was explained by turnover between schools. There were valid scores for 1,677 treatment school pupils, and 1,177 comparisons, with a total for the study of 2,854.

Figure 1: Recruitment and attrition
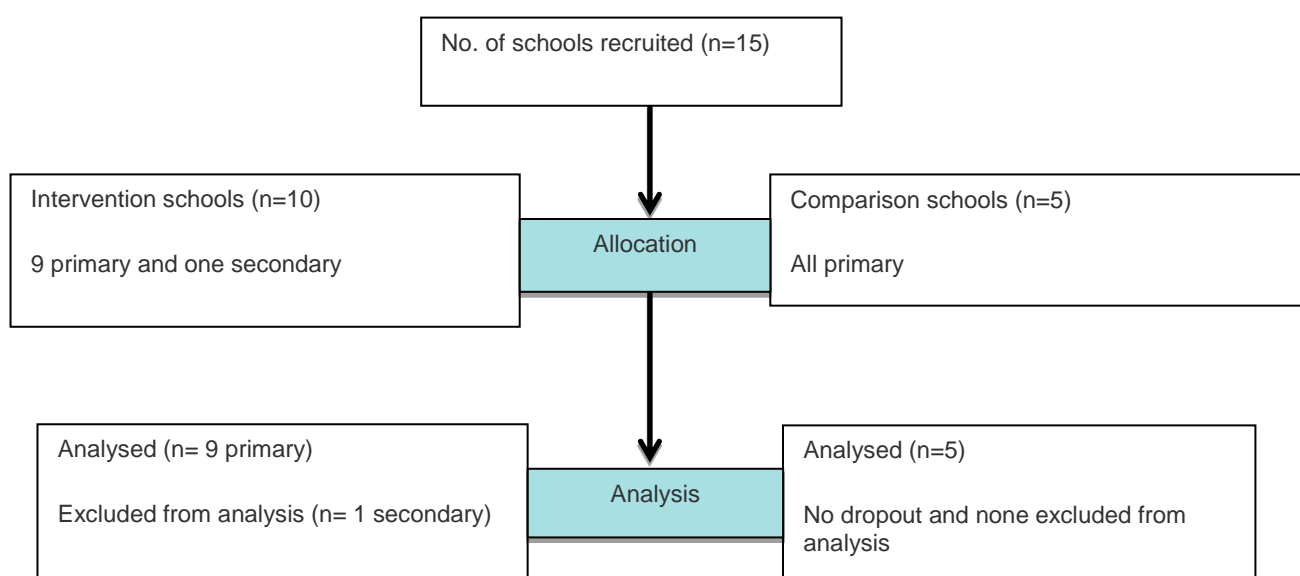
Note: This is not a trial, and there is no cluster randomisation or randomisation of any kind. All treatment schools are in one school partnership in Bexley. Comparison schools are those in the same area who agreed to provide data. Analysis is conducted by comparing pupils from the nine intervention schools with the five comparisons schools as well as with all schools in the Bexley area over two years.

Outcomes and analysis

Using all pupils with valid pre and post fine point scores in all years combined, the overall estimated impact of the intervention on pupil gain is negligible in all three subjects (Tables 1 to 3). There are very small negative 'effect' sizes in both reading and writing, and an equally small positive one in maths achievement.

Table 1. Effect size of gain scores for reading, all years combined

	N	Gain	Standard deviation	Effect size
Intervention	1676	4.16	2.97	-0.04
Comparison	1173	4.27	3.16	-
Overall	2849	4.20	3.05	-

Table 2. Effect size of gain scores for writing, all years combined

	N	Gain	Standard deviation	Effect size
Intervention	1649	3.95	2.82	-0.05
Comparison	1177	4.08	2.99	-
Overall	2826	4.01	2.89	-

Table 3. Effect size of gain scores for maths, all years combined

	N	Gain	Standard deviation	Effect size
Intervention	1677	4.17	4.17	+0.05
Comparison	1174	4.02	4.02	-
Overall	2851	4.11	4.11	-

Just under 12.6% of pupils are identifiable as being FSM-eligible. Of the 360, 199 are in the treatment schools and 161 in the others. Analysed separately, these yield larger and more consistent 'effect' sizes for the intervention, calculated in terms of gain scores (Table 4). These results do not have the authority of a trial, nor do they rule out a pre-existing difference between local schools and the Anglican Schools Partnership in terms of handling FSM pupils. But they do suggest a possibility that this intervention is especially important for FSM pupils, reducing the learning gap between them and others over one year, especially in maths. Converted into months' progress, effect sizes of +0.17, +0.12 and +0.41 are the equivalent of an additional learning gain of three, two and six months respectively during the course of a year.

Table 4. Effect size of gain scores for FSM-eligible pupils, all years combined

KS2 subject	'Effect' size
Reading	+0.17
Writing	+0.12
Maths	+0.41

N=360

It is arguable that the gains in fine point scores are not directly comparable across different years (especially as far removed as Years 2 and 6). Appendix B shows the results for each subject in each year from 2 to 6 (Tables 7 to 21). None of these results substantially alters the overall finding. However, it is noteworthy that the most substantial increases for the treatment schools occur in all

subjects in Year 4. It may be that this is a key year for enhanced feedback, more mature than Years 2 and 3, but still relatively unrestricted by the demands of preparation for KS2.

Table 5 presents the R values for three regression models, each based on two steps. Each model is used to try to explain variation in the gain score for each KS2 subject. In Step 1, the pupil background and prior attainment scores are included, and then in Step 2 the binary variable for being in the treatment group or control is added. For all three models the bulk of the variation that is 'explained' by the variables in the model is explained at step 1. Once pupil background and prior attainment is accounted for, very little difference is made by knowing whether a pupil was in the treatment group or not. This model is not, in itself, any test of causation but it does confirm the overall finding, and provides a caution about the strength and importance of the intervention in relation to prior pupil characteristics.

	Reading outcome score	Writing outcome score	Maths outcome score
Step 1: background and prior attainment	0.85	0.86	0.86
Step 2: treatment group	0.85	0.86	0.86

For completeness, Table 6 presents the coefficients for all variables retained in the three models. The largest of these by some way is the pre-test score. This is the best single predictor of the subsequent Key Stage attainment score, followed by having a reported special educational need and known eligibility for free school meals. Being in the treatment group has a very small positive standardised coefficient for all three subjects. It appears, at least, that no harm was done to the treatment pupils by trying out enhanced feedback.

Table 6. Standardised coefficients for the regression model in Table 5

	Reading outcome score	Writing outcome score	Maths outcome score
FSM	-0.04	-0.06	-0.04
Sex (female)	-	-0.02	+0.02
SEN	-0.07	-0.05	-0.04
EAL	-0.03	-0.04	-0.02
Ethnicity (non White UK)	-0.03	-0.02	-0.03
Prior attainment score	+0.82	+0.83	+0.84
Step 2: Treatment (or not)	+0.06	+0.01	+0.04

Table 7 shows two different kinds of comparators. Here the nine Anglican schools in the intervention are compared to the 49 other state-funded primary schools in Bexley. In 2012 before the intervention, 78% of the pupils in the treatment schools achieved Key Stage Level 4 or higher in reading, writing and maths. In 2013 after the intervention, 83% of the next cohort achieved what was deemed to be the same standard. All other schools in Bexley combined had a slightly lower percentage than the treatment schools in both years. But the overall difference in change is small, just as with the regression analysis and the direct comparison with the five comparator schools. Table 7 also includes a comparison of the average 2013 value-added scores for progress from KS1 to KS2 for the 422 Year 6 pupils in the treatment schools and the 2,187 Year 6 pupils in all other Bexley schools. Both groups

have a VA score that is indistinguishable from the mean of all primary schools in England (100). However, after these results are considered, there is no convincing evidence of a beneficial impact from this brief intervention.

Table 7. Comparison between intervention schools and all other primary schools in Bexley, progress 2012 to 2013, and value-added scores 2013

	Percentage achieving Level 4+ in reading, writing and maths, 2012	Percentage achieving Level 4+ in reading, writing and maths, 2013	KS1 to KS2 VA score, 2013
Treatment	78.2	83.0	100.0
Bexley	77.5	81.6	100.2
England (state primary)	75	74	-

Source: compiled from DfE School Performance Tables

Cost

The cost of running the pilot itself, including training of staffing and provision of resources is estimated at £39,000. The cost of the project in schools, including staff cover and conducting the progress surveys, is £49,000. With 10 schools (only 9 used as part of this evaluation) having around 4,000 pupils involved (including Year 1 and secondary pupils not part of the evaluation), the total cost per pupil is around £22. As conducted, this was a very cheap intervention.

Conclusions and implications

Limitations

This is a large-scale and in-depth study of Years 2 to 6 involving nine primary schools. The study took place in one authority and focussed on faith-based schools. The lesson observations were by prior agreement and so, as ever with such an approach, the evaluators cannot be sure that their presence had no effect on the lessons planned and conducted or the behaviour of the pupils. As a pilot study based on action research, the impact evaluation has no true counterfactual, and so the results must not be mistaken for those of a trial.

Interpretation

As explained in the introduction, despite large meta-analyses of mixed and sometimes poor studies, and the widespread acceptance that enhanced feedback is a good thing, there is no definitive evidence from trials that getting teachers to use more sophisticated and evidence-informed approaches to feedback will cause a rise in attainment. This new study does not change that, but it shows again the variability that is likely to result from a widespread push for more enhanced feedback. Teachers need more resources and examples from the outset, and cannot work from undigested evidence reports. There needs to be a clear and unbiased conduit from primary evidence to proposed classroom practice.

The impact evaluation suggests no overall difference between the schools using enhanced feedback and those carrying on with standard practice. This is the overall finding, whether based on a contrast with the designated comparator schools, value-added scores or progress of all other primary schools in the same local authority. However, it may be worth noting for some reason(s) the FSM-eligible pupils in the treatment schools improved disproportionately compared to the designated comparator schools, particularly in maths. If this kind of intervention is effective maybe it is more likely to assist FSM pupils than others (on average).

Future research and publications

If the promise of reducing the gap between FSM and other pupils glimpsed in these results is deemed worthwhile, the next step would be a formal trial ideally focused on FSM pupils only. If this step is taken, the advice from this pilot and process evaluation would be that the introduction of enhanced feedback should be made more structured from the outset. The action research approach adopted here appears genuinely useful for initial development, but future developments should focus on more codified intervention templates, with practitioners given more examples, resources and direction at the outset. Further development work is needed to make this feasible.

The evaluators are likely to produce a research paper based on these findings.

References

Black, P. and Wiliam, D. (1998). Inside the black box: raising standards through classroom assessment. London: GL Assessment.

Black, P., Harrison, C., Lee, C., Marshall, B. and Wiliam, D. (2004). Working inside the black box: assessment for learning in the classroom. Phi Delta Kappan, Vo. 86, pp. 8-21.

Fuchs, L.S. and Fuchs, D. (1986). Effects of Systematic Formative Evaluation: A Meta-Analysis, Exceptional Children, Vol. 53, pp. 199-208.

Gorard, S. (2006). Towards a judgement-based statistical analysis, British Journal of Sociology of Education, 27, 1, 67-80.

Gorard, S. and See, BH. (2013). Overcoming disadvantage in education, London: Routledge.

Hattie, J and Timperley, H. (2007). The power of feedback. Review of Educational Research, Vol. 77, No. 1, pp, 81-112.

House of Commons (2009). House of Commons fourth report on the National Curriculum, Vol. 1. London: The Stationery Office.

Smith, E. and Gorard, S. (2005). 'They don't give us our marks': the role of formative feedback in student progress, Assessment in Education, 12, 1, 21-38.

White, B. and Frederiksen, J. (1998). Inquiry, modelling and meta-cognition: making science accessible to all students. Cognition and Instruction, Vol. 16. No. 1, pp. 3-118.

Appendix A. Notes on instruments

General evaluation questions for all programmes (this can be adapted and tailored to specific projects)

Teacher interviews:

Background of teachers interviewed

1. Why did you decide to be involved in the programme (e.g. Summer school).
2. Or why did the school decide to be involved (if interviewing head teachers)
3. Has the programme changed your attitude towards teaching reading or use of feedback in the classroom?
4. Has it changed your method or style of teaching in the classroom?
5. What changes did you make to your teaching?
6. Are there any concerns regarding the implementation of the programme?
7. Were there any colleagues or students who were uncooperative or resistant to this programme? How was this dealt with?
8. What kind of assistance did you receive when you needed help?
9. Are there any aspects of the programme that you'd like to change or make it better?
10. What observations have you made about the impact of the programme on:
    a. Students
    b. Staff
    c. Parents?

The Pupil Learner Survey instrument, produced by the developer.

1. Welcome! This survey is to help us know what kind of learner you are.

First question is.... What is your name?

2. Is someone helping you to complete this survey?

   Yes

   No

3. What is your Unique Pupil Number? (Ask your teacher to tell you)

4. Are you FSM6? (Ask your teacher to tell you)

○ ○ Yes

○ ○ No

5. Are you a boy or a girl?

○ ○ Boy

○ ○ Girl

6. How much do you agree with this statement?

I am an excellent learner!		
Agree	Disagree	Not sure

Appendix B. Estimated impact results by year group

As discussed in the section of the report on impact evaluation, this appendix presents the results for each KS2 subject (reading, writing and maths) for each of Years 6 to 2 separately.

Tables 8 to 10 show the results for each subject in Year 6 alone. It is clear that the intervention had no beneficial impact in Year 6 and may well have been harmful in reading. Only 54 pupils in total were eligible for FSM.

Table 8. Effect size of gain scores for reading, Year 6

	N	Pre-score	Post-score	Gain	Standard deviation	Effect size
Intervention	316	26.89	30.45	3.56	2.54	-0.56
Comparison	192	24.90	29.57	4.69	5.36	-
Overall	508	26.13	30.12	3.99	3.89	-

Note: Effect size for FSM-eligible only +0.09.

Table 9. Effect size of gain scores for writing, Year 6

	N	Pre-score	Post-score	Gain	Standard deviation	Effect size
Intervention	316	24.92	29.24	4.32	2.59	-0.24
Comparison	193	23.20	28.40	5.20	4.84	-
Overall	509	24.27	28.92	4.65	3.63	-

Note: Effect size for FSM-eligible only -0.11.

Table 10. Effect size of gain scores for maths, Year 6

	N	Pre-score	Post-score	Gain	Standard deviation	Effect size
Intervention	314	26.10	30.06	3.93	2.33	-0.08
Comparison	191	24.49	28.90	4.35	5.19	-
Overall	505	25.49	29.62	4.09	3.68	-

Note: Effect size for FSM-eligible only +0.10.

Tables 11 to 13 show negligible differences between the two groups in Year 5, similar to the overall headline result. Only 68 pupils in total were eligible for FSM.

Table 11. Effect size of gain scores for reading, Year 5

	N	Pre-score	Post-score	Gain	Standard deviation	Effect size
Intervention	237	23.50	27.18	3.66	2.85	-0.06
Comparison	219	22.12	25.90	3.81	2.34	-
Overall	456	22.84	26.56	3.73	3.53	-

Note: Effect size for FSM-eligible only +0.01.

Table 12. Effect size of gain scores for writing, Year 5

	N	Pre-score	Post-score	Gain	Standard deviation	Effect size
Intervention	238	21.18	24.77	3.60	2.53	+0.06
Comparison	220	19.66	23.09	3.45	2.48	-
Overall	458	20.45	23.97	3.53	2.50	-

Note: Effect size for FSM-eligible only +0.21.

Table 13. Effect size of gain scores for maths, Year 5

	N	Pre-score	Post-score	Gain	Standard deviation	Effect size
Intervention	237	22.37	26.16	3.79	3.21	+0.01
Comparison	219	21.24	25.00	3.77	2.27	-
Overall	456	21.82	25.60	3.78	2.79	-

Note: Effect size for FSM-eligible only +0.14.

Tables 14 to 16 show a considerable difference between the two groups in Year 4, the mirror image of results in Year 6. Only 66 pupils in total were eligible for FSM.

Table 14. Effect size of gain scores for reading, Year 4

	N	Pre-score	Post-score	Gain	Standard deviation	Effect size
Intervention	324	20.43	24.61	4.19	1.86	+0.52
Comparison	223	18.99	22.19	3.19	1.84	-
Overall	547	19.84	23.62	3.75	1.79	-

Note: Effect size for FSM-eligible only +0.21.

Table 15. Effect size of gain scores for writing, Year 4

	N	Pre-score	Post-score	Gain	Standard deviation	Effect size
Intervention	324	20.43	22.24	3.86	1.90	+0.14
Comparison	224	18.99	20.39	3.60	1.62	-
Overall	548	19.84	21.49	3.75	1.79	-

Note: Effect size for FSM-eligible only +0.01.

Table 16. Effect size of gain scores for maths, Year 4

	N	Pre-score	Post-score	Gain	Standard deviation	Effect size
Intervention	324	19.64	23.78	4.14	2.06	+0.28
Comparison	224	18.30	21.89	3.59	1.77	-
Overall	548	19.09	23.01	3.91	1.97	-

Note: Effect size for FSM-eligible only +0.12.

Tables 17 to 19 show a mixture of apparent differences between the two groups in Year 3, with small positive outcomes in reading and maths. Only 67 pupils in total were eligible for FSM. However, the relative gains for FSM pupils were large and consistent across subjects. There are indications that despite the overall low impact of the intervention it may be more effective for FSM pupils.

Table 17. Effect size of gain scores for reading, Year 3

	N	Pre-score	Post-score	Gain	Standard deviation	Effect size
Intervention	397	17.03	20.90	3.86	2.31	+0.10
Comparison	223	16.66	20.09	3.49	2.35	-
Overall	620	16.90	20.61	3.73	2.33	-

Note: Effect size for FSM-eligible only +0.59.

Table 18. Effect size of gain scores for writing, Year 3

	N	Pre-score	Post-score	Gain	Standard deviation	Effect size
Intervention	398	15.60	18.72	3.12	2.00	-0.02
Comparison	223	15.03	18.16	3.16	1.90	-
Overall	621	15.40	18.52	3.13	1.97	-

Note: Effect size for FSM-eligible only +0.34.

Table 19. Effect size of gain scores for maths, Year 3

	N	Pre-score	Post-score	Gain	Standard deviation	Effect size
Intervention	397	16.52	19.51	2.97	2.13	+0.11
Comparison	224	16.50	19.23	2.75	2.05	-
Overall	621	16.51	19.41	2.90	2.11	-

Note: Effect size for FSM-eligible only +0.70.

Tables 20 to 22 show a mixture of apparent differences between the two groups in Year 2, with a negative outcome for reading, a neutral one for writing, and a negative one for maths. Only 105 pupils in total were eligible for FSM.

Table 20. Effect size of gain scores for reading, Year 2

	N	Pre-score	Post-score	Gain	Standard deviation	Effect size
Intervention	403	12.23	17.41	5.18	4.15	-0.13
Comparison	318	11.65	17.28	5.64	2.54	-
Overall	721	11.97	17.35	5.38	3.54	-

Note: Effect size for FSM-eligible only -0.23.

Table 21. Effect size of gain scores for writing, Year 2

	N	Pre-score	Post-score	Gain	Standard deviation	Effect size
Intervention	403	11.36	16.12	4.84	4.05	-0.00
Comparison	318	11.34	16.18	4.84	2.82	-
Overall	721	11.35	16.14	4.84	3.53	-

Note: Effect size for FSM-eligible only +0.07.

Table 22. Effect size of gain scores for maths, Year 2

	N	Pre-score	Post-score	Gain	Standard deviation	Effect size
Intervention	405	11.43	17.21	5.78	3.30	+0.21
Comparison	318	11.99	17.17	5.19	2.01	-
Overall	723	11.67	17.19	5.52	2.82	-

Note: Effect size for FSM-eligible only +0.32

Education
Endowment
Foundation

The Education Endowment Foundation

9th Floor, Millbank Tower

21–24 Millbank

London

SW1P 4QP

www.educationendowmentfoundation.org.uk