Teachers' use of research evidence in practice: a pilot study of feedback to enhance learning

Beng Huat See, Stephen Gorard and Nadia Siddiqui
School of Education, Durham University

Corresponding author
Beng Huat See
Email: b.h.see@durham.ac.uk
Tel: 07547048940

ABSTRACT

Background
There is worldwide interest in improving the effectiveness of teachers and teaching. This paper considers two strands of that interest. It revisits the impact of using enhanced feedback from teachers to pupils as a way of improving attainment, and it looks at the feasibility of teachers using research evidence to create their own interventions. Current evidence on the causal impact of effective feedback on learning is unclear: many studies have mixed results, are small in scale, lack randomisation or are not conducted in real classroom conditions.

Purpose
The aim of this paper is to describe the experience of schools as they engage with research evidence to support their own enquiry into the effectiveness of feedback in the classroom.

Research design
This study took place over one academic year, involving nine treatment schools in one local authority. The study involved teachers themselves using research findings to create an intervention, which took, as its focus, enhanced feedback in the classroom. Test results from these schools were compared to the results in five participating comparator schools, to the 49 other schools in the borough and to all state-funded primary schools in England.

Results
Although teachers showed that they could engage with research evidence, the study indicated that the process was complex in practice. In addition, the independent impact evaluation suggested that enhanced feedback in itself does not necessarily lead to improved pupil test performance.

Discussion and conclusions
The paper considers some of the challenges faced by teachers as they attempted to use research evidence, and discusses implications for schools wishing to use research evidence in practice. The findings of the study suggest that it may be feasible for practitioners to use research evidence to inform their own practice. However, to do it well would require clearer guidance, professional development and modelling of any strategies suggested. These findings have implications for policy on teacher development, and for the research community to make research outputs more comprehensible and accessible to research users.

(Number of words: 298)

KEYWORDS

Feedback, formative assessment, continuing professional development, teacher-led enquiry, teacher quality, action research

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors

INTRODUCTION

Aims

The aim of this paper is to describe the experience of schools as they engage with research evidence to support their own enquiry into the effectiveness of feedback in the classroom. It discusses the challenges they faced and the lessons learnt which may be of benefit to those who wish to embark on such a practice. It is not the primary aim of this paper to evaluate the impact of using enhanced feedback in the classroom, although the impact is also discussed briefly to consider whether the implementation of enhanced feedback in the classroom as interpreted by teachers was, in fact, effective in improving attainment. The paper concludes by looking at the implications of the findings.

This paper is, therefore, about teachers using research findings to create an intervention and modify it in light of experience. The intervention was enhanced teacher feedback to pupils. The main question addressed in this paper is:

> To what extent can teachers use research evidence to inform their classroom practice?

Given the considerable educational interest in the benefits of teacher feedback, a pertinent subsidiary question is whether the implementation of the intervention, as conceived by the teachers, led to improvement in pupils' attainment.

The paper starts by outlining current policy and discusses the need for teachers to conduct their own research as well as the evidence for using enhanced feedback in improving attainment. We then explain the methods used in the impact and process evaluations involved in the study. The paper continues with a discussion of how teachers engaged with research, presentation of the results of the impact evaluation, and some concluding suggestions.

Background

The use of research evidence in the classroom

Enhancing the quality of the teaching workforce is a key education reform policy in many developed countries (OECD 2011). It is widely acknowledged that the quality of teaching is an important school level factor influencing pupil achievement at school. In an effort to improve the quality of teaching, schools in the UK are now increasingly encouraged to use research evidence in the classroom to develop strategies to improve the learning outcomes for their pupils. This is in contrast with using personal experience and professional judgement alone or, worse, relying on commercially produced resources, many with no evidence of impact or even where the evidence is that they do not work (e.g. Khan and Gorard 2012).

In the UK, there have been attempts to facilitate the use of existing research evidence by teachers and school leaders by a number of initiatives. These include the Education Endowment Foundation pupil premium toolkits (http://educationendowmentfoundation.org.uk/toolkit/), which provide teachers and early-years practitioners with estimates of the impact, cost and security of a growing range of evaluated interventions. There are also a growing number of organisations led by teachers, such as ResearchED (http://www.workingoutwhatworks.com/en-GB/About)

which involve research leads, or 'champions' in participating schools promoting the use of research evidence in practice.

The impetus for evidence-informed practice has been fuelled by *Building Evidence into Education* (Goldacre 2013), which has spurred arguments for teacher-led enquiry. While there is evidence that some teachers in the UK are already using research to inform their practice, what is less clear is whether teachers are able to interpret that research easily, and whether it is possible for teachers to contribute effectively to the knowledge base by undertaking research for themselves. Among the many recommendations in the UK's House of Commons report on national curriculum reform (2008-2009) was the call for the then Department of Children, Schools, and Families (now called the Department for Education) to 'divert resources away from the production of guidance to the funding and dissemination of research findings to teachers in the spirit of informing local professional decision-making.' (House of Commons 2009, p. 41)

Many believe that teachers, as potential consumers of research, should be given the opportunity and the means not only to engage with the existing evidence, but also to conduct their own research to test the range of programmes they propose using in their classroom. A recent UK inquiry into the role of research in teacher education argued for teachers to be 'equipped to conduct their own research, individually and collectively, to investigate the impact of particular interventions or to explore the positive and negative effects of educational practice' (BERA-RSA 2014 p. 11). A review by the National Foundation for Educational Research (NFER) also argued for more evidence on the relative benefits of practitioner-led investigation.

There are indications that it is possible for schools and teachers to undertake robustly-designed research that adds to the accumulation of knowledge about the effectiveness of particular approaches (Gorard et al. 2015, Siddiqui et al. 2015). The evidence suggests that, in order to be able to carry out well designed research, there is a need for teachers to be trained and guided on important aspects of research craft, such as attention to bias and threats to validity and reliability.

The use of enhanced feedback to raise attainment

The teacher-led intervention in this study was based on a paper by Hattie and Timperley (2007). This paper is an important summary of evidence on the use of effective feedback, and, along with the equivalent section on feedback in Hattie (2008), has had considerable impact worldwide. However, this work does not, in itself, establish clearly that enhanced 'feedback' is generally effective. It is widely agreed that good teachers tend to use more and stronger feedback naturally. But what is not so clear is how easy it is to implement the use of such feedback by other, perhaps more reluctant, teachers, and whether this would make any difference to pupils' attainment.

In 1988, the National Curriculum Task Group on Assessment and Testing (TGAT) for England and Wales recommended that assessment should be an "integral part of the education process, continually providing both 'feedback' and 'feedforward' and ought therefore to be 'systematically incorporated into teaching strategies and practices at all levels" (DES 1988, paragraphs 3 and 4). The system advocated a combination of both formative and summative approaches. However, ensuring parity of esteem for both teacher-led formative assessment and summative assessment has been a contentious issue since the early days of TGAT. According to some commentators, this was motivated, at least in part, from 'political

concerns for accountability, rather than by educational concerns for learning' (Torrance and Pryor 1998, p. 10).

One strategy that has been advocated for use in the classroom is known as assessment for learning (AfL), a kind of formative assessment (defined as an activity that provides feedback to inform teaching and learning). A well-known UK review by Black and Wiliam (1998) called "Inside the Blackbox" indicated a substantial impact for AfL on learning for all age groups (from pre-school to undergraduate level), and claimed that formative assessment was especially effective for low achievers. Black (2000) cited research where the use of formative assessment techniques produced learning gains with effect sizes of between 0.4 and 0.7, larger than those produced by some other significant educational interventions. Fuchs and Fuchs (1986) reported benefits for children with special needs. White and Federiksen (1998) also reported that feedback was more effective for low achieving students (effect size of 1.0) than for high achieving students (effect size of 0.27).

More recently, Truckenmiller et al. (2014) conducted a small trial of feedback and found a 0.66 effect size, while Lipko-Speed (2014) confirmed in a small study that use of feedback was better than using further study or testing only. The early Education Endowment Foundation toolkit of interventions referred to above suggested that enhanced feedback had the potential to improve pupil attainment with an indicative effect size of 0.6 (Education Endowment Foundation 2015). Enhanced feedback, therefore, may hold considerable promise – arguably more than many other tested educational interventions.

There are two main problems with this array of evidence. Some of it is not particularly robust, and little sheds light on what happens when feedback practices are implemented across a wide variety of schools and teachers. The White and Federiksen (1998) study was based on only three teachers in two schools. Truckenmiller et al. (2014) had only 39 cases in the smallest group, and Lipko-Speed (2014) involved only 65 5th grade students (10-11 year-olds) altogether. The original Black Box experiment involved six volunteer schools, and it is evident that the approach does not always lead to equivalent success in less propitious circumstances (e.g. Smith and Gorard 2005). Other studies are not directly about the impact of enhanced feedback on attainment. For example, some are about how possible it is to implement an intervention such as AfL (Jonsson et al. 2015), while others are about the nature and perception of the feedback generated by teachers (Carvalho et al. 2014).

A newer review by Hopfenbeck and Stobart (2015) identified 1,387 reports of research on AfL. Notably, most of those concerned with impact on pupil attainment were small case studies (with perhaps one or two schools), while very few were large-scale or well-designed evaluations. This raises the question of which studies were synthesised by Hattie and Timperley (2007) – and it must be recalled that the latter was a synthesis of meta-analyses, themselves each a synthesis of very many studies.

Hattie and Timperley (2007) reported wide variations in effect sizes among studies, probably depending on the type of feedback used. In this, they confirmed the earlier findings of Kluger and DeNisi (1998) that while feedback interventions did improve pupil performance on average, many did not. Of the 607 effect sizes quoted, over one third were negative and associated with decreased performance. This cannot be explained by scale (sampling error) or even by theories of feedback use. But an explanation might involve the kinds of feedback and the nature of the pupils or students involved.

What these meta-meta-analyses appear to do is aggregate scores from a wide variety of research designs, practical approaches and educational settings. Some studies were specifically for children with special educational needs, or children with behavioural needs, emotional needs and disruptive behaviour. Some were based on very young children, and some used undergraduates at university. The individual studies used different calculations of 'effect' sizes, and often for different measures of the same parameters (e.g. different types of reinforcement and a range of feedbacks). The meta-analyses take no account of the bias introduced by attrition, treating a study with full response as equivalent to one with high dropout or missing data. Some include both passive designs along with randomised control trials (RCTs), some only the former, and no distinction seems to be made between them.

Hattie and Timperley (2007) describes a meta-analysis that included 196 studies and 6972 effect sizes, where the average effect size for feedback was 0.79 (Hattie and Timperley 2007, Table 2 p. 84). However, even a preliminary consideration of the cited evidence begins to suggest that such mega-syntheses may not be entirely accurate. For example, the review states that the effect size of 54 studies in Lysakowski and Walberg (1982) was +1.13 whereas the original paper reports it at 0.97 (study-weighted). However, the figure 1.13 does not appear in their paper. In Hattie (1992) and repeated subsequently, it is said that "Skiba, Casey & Center (1986) used 315 effect-sizes (35 studies) to investigate the effects of some form of reinforcement or feedback and found an effect-size of 1.88", but the later 2007 paper reports this review as having 35 effect sizes not studies, and an effect size of +1.24. While none of this undoes the work that has been done or eliminates the evidence for the impact of enhanced feedback, it ought to lead to caution. Overall, we conclude that the evidence is not as clear as some commentators have suggested.

The second problem is that few of the studies come with a practical guide on how to apply the research evidence in real classroom situations. For example, Kluger and DeNisi's meta-analysis was based on studies where the research was undertaken largely in controlled or laboratory conditions. Results may be different in real classroom conditions. Others have suggested that feedback may not always be effective if introduced without proper training for implementing it (Black and Wiliam 1998). Black and Wiliam stressed that for the successful implementation of formative assessment, pupils needed to be trained to: set goal oriented criteria (what Hattie termed 'Success Criteria' (Hattie and Timperley 2007, p. 88), assess their own progress, identify areas that need improvement and understand strategies needed to achieve this. White and Frederiksen (1998), on the other hand, suggested that pupils should be told explicitly by teachers what the criteria of success will be.

The feasibility of implementing the "Hattie model" of feedback strategy across year groups in primary schools has not been tested on a large scale in normal classroom conditions. It is against this background that a group of schools in a suburb of London decided to apply for funding from the Education Endowment Foundation to develop the feedback strategies advocated in Hattie and Timperley's meta-analysis and to trial this intervention for themselves.

METHODS

The project was a quasi-experiment, involving a feedback intervention. Although the intervention was devised and conducted by the schools themselves, it was independently evaluated by the authors, who monitored and observed the process by which the schools developed feedback strategies. The intervention was carried out in nine primary schools and one secondary school forming the Anglican School Partnership in Bexley, England, and

involving all Year 2 to Year 6 pupils (pupil ages 6 to 11) over one academic year. The single secondary school was not included in the analysis here because it had no participating comparator. The Partnership schools recruited five further local primary schools to provide data and so act as a comparison group. In addition, the evaluators compared the Year 6 pupils (10 -11 year-olds) in the treatment group to the performance of the 49 other state-funded primary schools in the borough, and to all such primary schools in England, using official statistics. A total of 1,677 pupils in Years 2 to 6 were involved in the intervention. A total of 2,187 Year 6 pupils in the 49 comparator schools, and 1,177 pupils in Years 2 to 6 in the five local schools, had assessment scores for comparison.

The schools themselves came together and decided to test this intervention in their own schools. All the teachers in the schools were involved in the project. The comparator schools were not matched in any way; they were simply similar schools in the borough which were not involved in the project.

This project was conducted in accordance with the British Educational Research Association's professional Code of Practice, and approved by Durham University's Ethics Committee. All participants in interviews and observations were informed that participation was voluntary and that they could withdraw consent at any stage. School visits were arranged with prior notice and with the co-operation of the school leads. Since the schools themselves conducted the research, the intervention was seen as something that the schools were doing anyway as part of normal school activity. Therefore there was no need for parental consent.

Process evaluation

This paper is mainly about the outcomes of the process evaluation. The process evaluation provided formative evidence on all phases of the intervention, from cascading the training to evaluating the outcomes. It involved participant observations and face-to-face interviews with teachers, project leads and pupils. Observations involved evaluators participating in the training, evaluation meetings and classroom delivery. These were as simple and as integrated as possible, in order to minimize disruption to lessons. The role of the authors as independent evaluators was simply to observe and monitor the programme as it progressed from one cycle to another. We selected a sample of four treatment schools for the process evaluation, based on their location for convenience of transport.

Altogether, the independent evaluators attended eight of the staff meetings. These included the initial training, review meetings (where staff came together to review progress and share their experiences) and moderation/training sessions. During these meetings staff shared their experiences and challenges faced in implementing the intervention in their schools. Ad hoc interviews with teachers and lead teachers were also carried out during these meetings. Additionally, evaluators visited four of the treatment schools to observe classroom delivery and implementation of the intervention. Visits were made at least twice, once at the beginning of the cycle and one towards the end in order to register progress. During these school visits, the evaluators observed sample classes from each of Key Stage 1 (Years 1 and 2; pupil ages 5 – 6 and 6 - 7) and Key Stage 2 (Years 3 to Year 6; pupil ages 7 – 8 to 10 - 11), and interviewed staff and pupils. All in all, 24 person visits were made to research sites.

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 the intervention was taking place. This meant teachers checking for effects of the action over and above what might otherwise have been expected, learning what seemed to work best and what the barriers were, and modifying the action for the next step in the cycle. The process evaluation looked specifically for data that addressed:

- the reaction to training
- the fidelity of training when it was cascaded down
- whether the teams understood the process and purpose
- the contents and use of the starter pack
- starting point and subsequent assessments
- changes in teacher behaviour
- how teachers took control of their own 'feedback' loop in improving evidence-informed practice
- audits by classes of teacher feedback, and learner effectiveness
- the ongoing modification of the starter pack
- whether there appeared to be an impact on how children are learning
- whether teachers were providing useful and better feedback
- whether pupils were responding.

Observations and other field notes, and interview transcripts, were analysed and synthesised according to these points.

Impact evaluation
Although the main aim of this paper was not about the impact of enhanced feedback, a brief discussion of the results is necessary to assess whether the intervention, as interpreted and applied by teachers, was as effective as the research evidence suggested.

For the Year 6 pupils (aged 10 -11 years) in the nine treatment schools who had national test scores, , attainment was compared to the 49 other state-funded primary schools in the borough, and to all state-funded primary schools in England, based on progress measured by national test scores, and to improvements since 2012 (i.e. the year preceding the intervention). Because the treatment and control schools were not randomly assigned or matched, it was necessary to compare the treatment schools with all other schools in the area, and nationally, to get a better picture of how well the treatment schools actually did in comparison. For other year groups it was necessary to use the end of year teacher assessments rather than official national test scores at age 11.

The impact of the intervention was measured by calculating the differences in gain scores from the prior national test scores to the subsequent national test scores, and presented in standardised form as Hedges' g 'effect' sizes. The results are presented overall, by years and subjects (reading, writing, maths). Sub-group analyses by gender and school meals eligibility (FSM) were also carried out to assess the impact for different groups of children.

Tests of significance were not used, and confidence intervals are not provided for the effect sizes, because the cases being compared were neither randomly selected nor allocated (Berk and Freedman 2001). They also take no account of sample quality or attrition (Lipsey et al. 2012), and are therefore irrelevant here and would be misleading if presented (Gorard 2015).

The intervention

The intervention used in the project was enhanced feedback, based on Hattie's model of feedback which is taken largely from his paper: The Power of Effective Feedback (Hattie and Timperley 2007). The schools wanted their teachers to use the paper as a basis for devising classroom practices.

In this model, the first step for teachers involved identifying the gap between what the learner knows and what they need to know, as well as how they could reach that goal, or target, using feedback strategies to guide them towards self-regulatory proficiency. In this project, the teachers got the pupils to set their own goals (i.e. 'Success Criteria'). Prior to this, pupils completed an online Learner Effectiveness survey to establish their learning styles.

Teachers then collected examples of the three types of feedback and the four levels, as described in Hattie's paper:

      1. Feed Up   *Where am I going (the goals)?*
      2. Feed Back   *How am I doing?*
      3. Feed Forward   *Where to next?*

According to the model, each of these feedback questions can be directed at four levels: (1) Personal (Self), (2) Task, (3) Process, and (4) Self-regulation. The aim is to guide pupils towards the level of self–regulation. The effectiveness of feedback is influenced by how well teachers use these levels. Feedback at the personal level is about giving feedback about the person. At the task level, the teacher gives feedback about the task, that is, whether the task is correct or not and may include directions to achieve the correct answer. At process level, the feedback is about the process, that is, the learning process needed to complete a task. Feedback can also be directed at self-regulation. This is about giving feedback on the skills needed at self-evaluation. The most effective feedback is one that moves pupils from task to processing and from processing to self-regulation. It is important to note that the research article was not a teaching manual, so few examples were given. Teachers had to make their own interpretation of what each level and process involved.

The project

The project studied the process of teachers applying research evidence on feedback in the classroom, using an Action Research model of implementation. It involved teachers coming together to review the research evidence on effective feedback and developing strategies to use them in the classroom, employing what they described as a cyclical action research design, but was more nearly an example of design-based study (Gorard 2013). The participants include all the teachers in the participating schools.

A pack of training materials (the starter pack) was prepared by the school project leads. Working in pairs on a school training day, schools initially received training using these packs. The training involved the school leads reading and discussing Hattie's paper.

A moderation meeting was convened with project leaders and school leads, to help them agree on examples of the types of feedback. School leads then delivered feedback moderation training to staff. This was followed by a moderation staff meeting to establish starting points. Learning teams established starting points by carrying out audits, including

pupils' baseline data using a Pupil Learner Effectiveness (PLE) questionnaire devised by the school leads. The PLE survey and a feedback survey were administered. All pupils completed the teacher-developed online survey of Pupil Learner Effectiveness, which was aimed at identifying pupils' starting points and their learning strategies.

The study led by the teachers involved four cycles of action research, Action Research Cycle 1 to 4. At the end of each cycle, there was a Project Lead evaluation meeting to share examples of good practice.

Action research cycle 1
Each school lead was given three days of temporary teacher cover so that participants could collect examples of the proposed three types of feedback (feedup, feedback, and feedforward) and the four levels for each type (personal, task, process and self-regulating) in the Hattie model. Teachers also identified 'Where pupils are going', 'How they are doing' and 'Where to next'. Teachers audited each other's lessons to assess the prevalence of the 12 combinations of feedback. They then created an action plan aiming for a new balance of feedback, making it 'proportionate' to its value, looking for more self-regulatory feedback and fewer personal comments.

Action research cycles 2 and 3
Cycle 1 was followed by three further cycles, with the aim of moving pupils from having, for instance, the characteristics of ineffective learners (e.g. not planning) to effective (e.g. planning) learners. Essentially, an effective learner was considered to be one who knows where they are going and how they are doing and what they need to do to go to where they are aiming at. Therefore, students start each process by setting their own success criteria. They need to be able to identify their own mistakes and know what to do to correct them. This was the self-regulatory stage.

For cycles 2 and 3, each teacher audited pupils' skills in terms of the Hattie and Timperley (2007) model of being an effective learner using the results of the Pupil Learner Effectiveness survey. Teachers then reflected on their current practice and monitored pupils' understanding and application of the concept of an effective learner. The schools identified areas for improvement and strategies to achieve these. School leads then met to discuss issues and challenges and report on progress made. A second Pupil Learner Effectiveness survey was conducted at the end of the third action research cycle.

Action research cycle 4
This was the last cycle where teachers came together to analyse the post- Pupil Learner Effectiveness survey results and determine their pupils' progress as independent learners. The results are discussed later in the paper.

FINDINGS

To what extent can teachers use research evidence to inform their classroom practice?

The study found that although the approach of getting teachers to use research evidence to inform classroom strategies was feasible, the process was challenging and complex. Although schools were keen to show us their use of feedback in their lessons, it was not possible to tell if such attempts were used consistently or only for the observed lessons. There was also no way of telling whether the more successful teachers (in terms of their use

of the enhanced feedback) were already practicing it or whether the intervention made a difference to their pedagogical skills.

Although the majority of participants were very enthusiastic about the intervention, comments suggested some initial reservations. For example, some thought that the feedback strategy suggested in the model was something that they were already doing, or at least that good teachers should be doing, as these participant quotations demonstrate:

> *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?*
>
> *I do agree with him about the impact of feedback, but this is what we all do in our classes. This is what we are doing already.*

Perhaps more importantly, even the research leads struggled to understand the academic paper which set out the principles of effective feedback and distinguished between different types of feedback. It was difficult for the practitioners to engage with the paper, as it did not provide sufficient examples to explain the concept. Teachers were coming up with their own understanding and interpretation. These participant comments indicate some of the difficulties:

> *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'?*

We suggest that part of the problem is that the paper is not written in an easy to read way, and is therefore not accessible to the widest audience. This is not an issue of technicalities or of complex science, but of poor or overly complex writing (Gorard 2013). It highlights the need for a conduit between research reports, as published, and teachers' use of evidence (as previously proposed by Gorard with Taylor (2004).

In discussions, teachers were using the term 'feedback' as they understood it, but not the types espoused by Hattie and Timperley. For example, the literature on feedback draws an essential distinction between feedback targeted at the self and feedback which promotes self-regulation and independent learning. Examples of the latter given in the paper (Hattie and Timperley 2007, p. 90) would be '*You are a great student*' and '*That's an intelligent response, well done.*' An example of self-regulation would be: '*You already know the key features of an argument. Check to see whether you have incorporated them in your first paragraph.*' However, it was apparent from the observed lessons that this distinction was not consistently understood.

Implementation of feedback in the classroom

The intervention involved teaching pupils to set their own 'Success Criteria' (SC) for each session. Through appropriate use of feedback and directional strategies, teachers were to guide pupils to identify and correct their own mistakes. In this way, pupils would learn to develop strategies to achieve their learning goals. The aim was to move pupils to a self-regulatory stage as independent learners (hence the importance of setting their own Success Criteria). However, one difficulty was that such procedures were not consistently carried out, because not all teachers fully comprehended what Success Criteria should look like or what the different feedback strategies were.

Although there were some excellent innovative and creative lessons, where teachers attempted to encourage independent learning by modelling to pupils how to correct their own mistakes, the majority of lessons observed were evaluated as poorer in terms of feedback implementation. There was a lot of talk about feedback, but there were lessons where teachers simply got pupils to self-evaluate or peer-evaluate. Such evaluation was problematic, especially for the weaker pupils. Often, pupils did not know what was right or wrong about their work or their peer's work. School leads reported that pupils tended to be very polite and only said nice things about their friend's work. There was little focus on learning outcomes.

There were instances where pupils were asked to make their own Success Criteria but it was evident that some either had not done it or did not know how to. Nevertheless, 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 Success Criteria . There was no reciprocal feedback from pupils to teachers. The teacher was often not aware that pupils had not completed their initial Success Criteria forms. 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. Very often, teachers simply asked pupils if they had understood, rather than actually testing whether they had.

There was also no clear evidence that teachers were using the different levels of feedback – either orally or written. As time went on, teachers were using Success Criteria and lesson objectives alone, rather than feedback as such. The younger pupils did not understand what they were expected to do. One teacher's interpretation of Success Criteria was evident when they told their pupils that: "To be successful we will need to talk to our partners, work sensibly and share ideas".

There were also different interpretations of the processes and levels of feedback strategies across schools. Much of the feedback used in the classroom was at task and self level, and occasionally process level. There was hardly any feedup or feedforward. Generally, in a number of lessons observed, feedback was very limited and praise was used instead, such as: "Well done", "Excellent", "Brilliant". In one observation, throughout the lesson the teacher was making comments including:

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.

It is interesting to note that some teachers in training disagreed with Hattie about the likely negative effect of praise at self-level, and perhaps this is why they continued with it, despite being taught not to. Although such feedback was identified in the model as ineffective and in some instances even harmful, teachers were still using it a lot in the classroom. 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 specific aspects of learning. It was couched in vague terms such as, "lovely", "good" and "great". Because teachers were not specific in their feedback, pupils did not know what was good and what they needed to do to improve. School leads also reported that feedback in a number of lessons was generally in the form of praise and this was often not constructed in terms of what the next step should be for the pupil.

Although there were attempts to move pupils to self-regulatory feedback, 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 when checking. More guidance would be needed to steer pupils to meet their Success Criteria. There was generally still a lot of 'teacher talk' in every classroom.

The lack of actual examples illustrating the different levels and processes of feedback proved to be a big hindrance. As mentioned previously, the paper which teachers used to guide their practice did not provide such examples. The school leads did not supplement their resources from the wider literature. Their aim was for teachers to collect their own examples of feedback strategies and develop the resources to share among the schools. This was fraught with difficulties, as teachers were themselves unclear about what 'feedback', 'feed forward' and 'feedup' should look like.

There was also a lack of differentiation for the benefit of the more and less effective learners, which meant that the more able pupils may have 'ceilinged' – in other words, these pupils were not being challenged and stretched. In a number of mixed ability lessons observed, the less able pupils seemed to get more attention from the teaching assistants and teachers, while the more able received considerably less attention. Where ability grouping was used, this meant that there was limited opportunity for the less able pupils to model the effective learning strategy of the more able pupils. Pupils seemed very aware of where they were in terms of ability just by virtue of the colour of their tables. For example, those seated on the green table knew they were not the higher achievers, whereas those on the red table were generally left to get on with their own work independently.

What was the estimated impact on pupil outcomes?

In terms of gains for Year 6 (10 -11 year old) pupils only (the only year group where such data was available) from the end of the school phase Key Stage 1 (pupil ages 6 to 7) to the end of the school phase Key Stage 2 (pupil ages 10 to 11), results show that intervention pupils in the nine Partnership schools made slightly bigger gains in reading, maths and writing compared with their nearest neighbours, with an effect size of +0.04 and also when

compared with the 49 other state-funded primary schools (ES= +0.06). However, it is important to note that using the value-added scores, intervention pupils did not show much progress when compared to the schools in the Bexley area. Table 1 below shows the results of the comparisons.

Table 1 - Comparison between Year 6 pupils (10 – 11 year olds) in intervention schools and all other primary schools in Bexley, progress 2012 to 2013, and value-added (VA) scores 2013

	Percentage Level 4+ in reading, writing and maths, 2012	Percentage Level 4+ in reading, writing and maths, 2013	Key stage 1 to Key stage 2 VA score, 2013
Treatment schools	78.2	83.0	100.0
Bexley (other schools)	77.5	81.6	100.2
England (state primary)	75	74	-

Source: compiled from DfE School Performance Tables

Comparing the nine Partnership schools with the five neighbouring comparator schools, analysis for all year groups combined again showed small and mixed results (Tables 2 to 4). Intervention pupils made less gains in teacher assessment for reading (effect size of -0.04) and writing (-0.05) but slightly higher for maths (+0.05).

Table 2 – 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 3 – 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 4 – 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	-

Overall, the data indicate that there is no convincing evidence of a beneficial impact on pupil outcomes from this intervention, although sub-group analysis shows that the intervention may be particularly beneficial for FSM-eligible pupils. For example, the gain score 'effect' sizes were positive for FSM-eligible pupils in all three subjects - reading (+0.17), writing (+0.12) and maths (+0.41). However, the number of cases was small (360), and volatility and differences in reporting cannot be ruled out. On the other hand, it may be the case that the younger pupils needed more guidance and were not ready for independent learning, while

such an intervention may be more beneficial for disadvantaged older children, who are more ready for independent learning but at the same time not restricted by the demands of preparation for national assessments at the age of 11.

DISCUSSION AND IMPLICATIONS

What has been learnt about the intervention?

The impact evaluation shows that the intervention yielded mixed results. There is some indication that it may be beneficial for disadvantaged pupils. However, because of the small number of cases, the lack of randomization and partly because the outcomes for years other than Year 6 were teacher assessed and not blinded, the results must be treated with caution. It must be borne in mind that this was a pilot study, and that it is also possible that the lack of impact is attributable to poor implementation, despite the fact that those running the project, and the clear majority of school leads and teachers, were very enthusiastic about the intervention. This study must therefore add to the cautionary literature, not so much about the ineffectiveness of feedback itself but about the difficulties of taking academic research evidence such as that generated in controlled conditions and implementing it more widely in everyday classroom conditions.

Although the use of feedback may seem like a natural part of practitioner activity, since it is routinely used in the classroom, for all teachers to be able to use it effectively requires guidance and practice. One of the aims of this project was to encourage teachers to use higher levels of feedback to encourage learning. What this study has indicated is that some teachers do not recognise what these higher levels of feedback look like, and are, thus, unable to use them to enhance learning. Nevertheless, the experience from this project has provided some valuable lessons for schools which wish to translate theory into practice. We suggest that some key implications for practice are as follows:

- Practitioners need to be given more examples, resources and direction at the outset of any change in practice. In the case of feedback, ample examples of 'Success Criteria', and different types of feedback should be made available in a resource pack or on a website which teachers can have easy access to. Video recordings of effective lessons could be used as a training resource so that teachers can model these lessons.

- Teachers need to be clear about what 'Success Criteria' are, what the different processes and levels of feedback look like and how to use feedback strategies to guide pupils to achieve their Success Criteria. Success Criteria need to be phrased in specific or measurable terms so that pupils know when they have achieved them or not.

- There should be a conscious effort to use higher levels of feedback to guide pupils to self-regulation, and minimise the use of self-feedback, which previous research has suggested to be the least effective – such as 'you are a superstar' and 'This is a clever idea'.

- All feedback should be clear, simple, specific and directed appropriately to specific learners.

- Teachers need to make appropriate judgements about when, how and what level of feedback is suitable for pupils.

- Greater differentiation in the use of feedback is needed, including more use of 'feed up' and 'feed forward' for the more able pupils, to provide the challenge.

- Effective classroom instruction must be used in concert with feedback. Feedback should not be a substitute for classroom instruction, nor for teacher content-pedagogical and assessment knowledge (Herman et al. 2015). For example, telling pupils that they need to use more interesting vocabulary is not helpful if pupils have not learnt the vocabulary. Neither is saying to pupils, "check your answers again", if pupils cannot see what is wrong with their answers. Hattie and Timperley (2007) 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 surely of little use.

What has been learnt about teachers' use of evidence?

The study shows that, although teachers can engage with research evidence, the process is complex in practice. This is partly because academic papers are not written for practical application and not specifically meant for practitioners. Such papers do not usually give detailed descriptions of interventions and how they are to be implemented. This is especially true of meta-analyses and syntheses of evidence, which are compilations of many different studies using a variety of measures and involving a range of outcomes for participants of different ages. For practitioners to use such evidence therefore requires more than simply reading the papers. We argue that there needs to be a clear and unbiased conduit from primary evidence to proposed classroom practice (Nelson and O'Beirne 2014). Teachers need relevant resources and examples from the outset.

These findings have implications for policy on teacher development. In their response to the Education Select Committee (House of Commons 2012), the UK government expressed the wish to encourage teachers to engage in and with research. There is, therefore, more to be done to support teachers as researchers and reflective practitioners. There was a suggestion in the government response to the House of Commons report on the national curriculum reform (House of Commons 2008-2009, paragraph 92) that resources, including a bank of pedagogical evidence would be built to support and engage teachers in developing pedagogy in the classroom. However, it remains unclear to what extent teachers have used or contributed to the resources or if the impact of these efforts has been independently and robustly evaluated.

Funding, resourcing and access to research evidence remain areas of concern. Currently, there are teaching scholarships in the UK for teachers to use towards CPD (Continuous Professional Development), and schools can apply for funding as well to do this. However, this can be a deterrent because of the time needed to apply for such funding. In addition, the cost of research-derived resources, some of which are only available through commercial providers, can also be a barrier to access.

The purpose of educational research is largely to inform policy and practice. A high proportion of education research is either publicly funded or funded by charitable organisations, and therefore it should follow that the outputs of such research should be made available to the people who want to consume (and help fund) the outcomes of the

research. This not only saves time and money for schools and teachers, but also enhances the professional capacity of teachers for the benefit of learners. We argue that it is the ethical responsibility of academics to make their research output as comprehensible as possible to consumers of research (in this case, the teachers). With the move towards open access in academic publishing, it is our hope that more of the academic papers that are relevant to classroom practice will become readily open to inspection by practitioners and the public. However, this also means that it is even more important for teachers need to have access to training and development, in order to judge the quality of evidence and be equipped with the necessary research skills to test such evidence for themselves.

REFERENCES

BERA-RSA(2014) *The role of research in teacher education: Reviewing the evidence. Interim report of the BERA-RSA inquiry*, London: British Educational Research Association

Berk, R. and Freedman, D. (2001) *Statistical assumptions as empirical commitments*, http://www.stat.berkeley.edu/~census/berk2.pdf, accessed 030714

Black, P. (2000) *Research and the Development of Educational Assessment,* Oxford Review of Education, 26, 3&4, 407-419.

Black, P. and Wiliam, D. (2003) In praise of educational research: formative assessment, *British Educational Research Journal*, 29, 5, 623-638.

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

Carvalho, C., Santos, J., Conboy, J. and Martins, D. (2014) Teachers' Feedback: Exploring Differences in Students' Perception, *Procedia - Social and Behavioral Sciences*, 159, 169-173

DES (1988) *National Curriculum: Task Group on Assessment and Testing: A Report,* London: DES/Welsh Office.

Education Endowment Foundation (2015) *Education Endowment Foundation pupil premium toolkit*, http://educationendowmentfoundation.org.uk/uploads/pdf/Feedback_Toolkit_references.pdf

Goldacre, B. (2013) *Building evidence into education*. London: Department for Education, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/193913/Building_evidence_into_education.pdf

Gorard, S. (2013) *Research Design: Robust approaches for the social sciences*, London: SAGE

Gorard, S. (2015) Rethinking "quantitative" methods and the development of new researchers, *Review of Education*, 3, 1, 72-96, doi: 10.1002/rev3.3041

Gorard, S., Siddiqui, N. and See, BH (2015) *Fresh Start: an aggregated trial*, Evaluation Report and Executive Summary, EEF, http://educationendowmentfoundation.org.uk/projects/fresh-start/

Gorard, S., with Taylor, C. (2004) *Combining methods in educational and social research,* London: Open University Press Hattie, J. (1992) What works in Special Education Presentation to the Special Education Conference, May, 1992, http://www.education.auckland.ac.nz/webdav/site/education/shared/hattie/docs/special-education.pdf

Hattie, J. (2008) *Visible Learning*, London: Routledge

Hattie, J, and Timperley, H. (2007) The power of feedback, *Review of Educational Research*, 77, 1, 81-112

Herman, J., Osmundson, E., Dai, Y., Ringstaff, C. and Timms, M. (2015) Investigating the dynamics of formative assessment: relationships between teacher knowledge,

assessment practice and learning, *Assessment in Education*, DOI:10.1080/0969594X.2015.1006521

Hopfenbeck, T. and Stobart, G. (2015) Large scale implementation of assessment for learning, *Assessment in Education*, 22, 1, 1-2

House of Commons (2009) *National Curriculum: Fourth report of session 2008-2009,* London: The Stationery Office Limited

House of Commons (2012) *Great teachers: attracting, training and retaining the best. Government response to the committee's ninth report of session 2010-2012. First special report of session 2012-2013*, London: The Stationery Office Limited

Jonsson, A., Lundahl, C.and Holmgren, A. (2015) Evaluating a large-scale implementation of Assessment for Learning in Sweden, *Assessment in Education*, 22, 1, 104-121

Khan, M. and Gorard, S. (2012) A randomised controlled trial of the use of a piece of commercial software for the acquisition of reading skills, *Educational Review*, 64, 1, 21-36

Kluger, A. and DeNisi, A. (1996) The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory, *Psychological Bulletin,* 119, 2, 254-284

Kluger, A., and DeNisi, A. (1998). Feedback interventions: Towards the understanding of a double-edge sword. *Current Directions in Psychological Science*, *7*, 67–72

Lipko-Speed, A., Dunlosky, J. and Rawson, K. (2014) Does testing with feedback help grade-school children learn key concepts in science*?, Journal of Applied Research in Memory and Cognition*, 3, 3, 171-176

Lipsey, M., Puzio, K., Yun, C., Hebert, M., Steinka-Fry, K., Cole, M., Roberts, M., Anthony, K. and Busick, M. (2012) *Translating the statistical representation of the effects of education interventions into more readily interpretable forms*, Washington DC: Institute of Education Sciences

Lysakowski, R., and Walberg, H. (1982) Instructional effects of cues, participation, and corrective feedback: A quantitative synthesis. *American Educational Research Journal*, *19*, 559–578

Nelson, J. and O'Beirne, C. (2014) *Using evidence in the classroom: What works and why?* Upton Park, Slough: National Foundation for Educational Research

OECD (2011) *Building a High-Quality Teaching Profession: lessons from around the World*, http://www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/ buildingahigh-qualityteachingprofessionlessonsfromaroundtheworld.htm

Siddiqui, N., Gorard, S. and See, BH (2015) Accelerated Reader as a literacy catch-up intervention during the primary to secondary school transition phase, *Educational Review,* http://www.tandfonline.com/doi/full/10.1080/00131911.2015.1067883#.Vbs-W0_bKUk

Skiba, R., Casey, A., and Center, B. (1985–1986) Nonaversive procedures in the treatment of classroom behavior problems, *Journal of Special Education*, *19*, 459–481

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

Torrance, H. and Pryor, J. (1998) *Investigating Formative Assessment: Teaching, Learning and Assessment in the Classroom,* Buckingham: Open University Press

Truckenmiller, A., Eckert, T., Codding, R. and Petscher, Y. (2014) Evaluating the impact of feedback on elementary aged students' fluency growth in written expression: A randomized controlled trial, *Journal of School Psychology*, 52, 6, 531-548

White, B. and Frederiksen, J. (1998) Inquiry, modelling and meta-cognition: making science accessible to all students. *Cognition and Instruction*, 16, 1, 3-118