



Institute for
Effective Education
Empowering educators with evidence

Audio feedback

Evaluating the impact of feedback using an audio tool compared to written feedback

John Coats

Notre Dame High School

February 2019



About IEE Innovation Evaluation Grants

The first four IEE Innovation Evaluation Grants were awarded in February 2017. Funded by the Institute for Effective Education (IEE), these grants supported pilot evaluations of innovations of teaching and learning approaches based on the Research Schools Network's goal of improving the attainment of pupils by increasing the use of evidence-based practices.

Since then a further 26 projects have been successful in their application for an IEE Innovation Evaluation Grant, bringing the total number to 30. The applications we received included a wide range of interesting, school-led innovations – from after-school film clubs to improve the creative writing of Year 5 pupils, to the use of audio feedback with Year 12 pupils – and we were really impressed with the thought that applicants had put into how these innovations could be evaluated.

The evaluations are small-scale, and test the kinds of innovations that schools are interested in. This is very much a “bottom-up” exercise, allowing schools to get some indicative evidence behind real-world initiatives. Many evaluations are now coming to an end, and we are starting to publish reports on the findings. It is important remember that these are small-scale projects, often carried out in one school, so it is not possible to generalise their findings. In fact, the main benefit of the Innovation Evaluation projects may be in the process, rather than the findings.

Contents

Section	Page
Executive summary	4
Introduction	5
Description of the problem	5
Review of existing research	5
Description of the innovation	6
Research questions	6
Method	7
Sample	7
Assignment to condition	8
Innovation	8
Outcome measures	10
Process evaluation	11
Analyses	11
Cost	12
Results	14
Outcome findings for primary research question	14
Outcome findings for secondary research question	14
Process evaluation findings	16
Discussion/Conclusion	23
References	25

Executive summary

Description of the innovation

Teachers of Year 12 A-level mathematics and sociology trialled recording digital audio feedback and sharing this with pupils (for example via school email) instead of providing more traditional written feedback. Pupils accessed the feedback through mobile devices, and where this wasn't possible, through home computers.

Summary of the evaluation

The study involved a total of 19 classes of Year 12 pupils from seven urban secondary schools where there was a lower than national average proportion of disadvantaged pupils. Assignment to treatment was carried out at a whole class level using GCSE Attainment 8 scores to minimise the difference in prior attainment between control and treated cohorts.

During the trial, pupils in treated groups received recorded audio feedback on three pieces of their work, while control groups received conventional written feedback.

The effect size was measured by comparing GCSE grades with post-test scores.

Summary of findings

The study found that the use of verbal feedback (using an audio tool), delivered over a discrete topic, was more effective than written feedback in improving test outcomes in sociology and mathematics A-level (effect sizes +0.15 for sociology and +0.18 for mathematics). Effect sizes were similar for boys and girls, but were negative (-0.64) for disadvantaged sociology pupils (although the number of these pupils was small).

We also found that the use of this intervention may have a positive impact on workload for teachers of A-level sociology, but that it had a statistically significant negative impact on the workload of mathematics A-level teachers compared to the work involved in providing conventional written feedback.

Mathematics pupils reported finding conventional written feedback more useful than the use of an audio tool (at a statistically significant level), whereas sociology pupils found the opposite (again at a statistically different level).

Introduction

Description of the problem

Teachers spend a long time providing written feedback to pupils.

There is limited evidence as to what type of feedback is most effective in moving pupils forward. Improvements in technology, development of apps such as Kaizena and widespread ownership of mobile devices means the recording of audio feedback is now a practical and viable alternative to providing written feedback. The effectiveness of this alternative form of feedback should be evaluated.

Teachers spend a significant proportion of their time marking work and workload is a major reason cited for current recruitment and retention issues (DfE, 2015). Evaluating the impact of alternative forms of feedback on teacher workload is therefore important for the entire profession.

Review of existing research

We know that feedback is a high-impact intervention when carried out well. Teachers traditionally spend large amounts of time marking individual pupil work as part of the process of providing feedback to pupils, but there isn't a large volume of quality research to help guide teachers as to how to use the time they are spending marking work most effectively.

“Despite its centrality to the work of schools and teachers, there is in fact little high-quality research related to marking.”

“There is an urgent need for more studies so that teachers have better information about the most effective marking approaches.”

(Elliott, et al., 2016)

With the advent of new technologies there are opportunities to explore the use of different ways of providing feedback to individual pupils. Hattie and Timperley (2007) cite a meta-analysis that suggests the effect size of feedback is greatest when the feedback gives the learner cues and is video or audio in format. This meta-analysis was carried out in 1999 before easy and efficient access to mobile technology.

One known barrier for pupils is understanding the terms used in written feedback. Elliott et al (2016) cite a study that found that university students often did not understand the terms used in written feedback. This resonates with our own anecdotal evidence from an informal pilot study where key stage 5 (KS5) teachers reported that their pupils were better able to understand audio feedback. There is also reference within *A marked improvement?* (Elliott, et al., 2016) to wider evidence concluding that the specificity of the feedback has an impact on performance. Our own experience, again gathered through our informal pilot study, is of teachers being able to give more detailed (ie, more specific) feedback in a given period of time via verbal (audio) feedback than via more traditional (written) feedback.

Our research will investigate whether audio feedback given by a teacher in the context of a busy working schedule is more effective than written feedback. That is, do the findings of Hattie and Timperley translate to 'less than perfect' conditions, where teachers have limited time in which to provide feedback (written or audio)?

Pilot

Prior to carrying out this innovation project, teachers of sociology and mathematics at Notre Dame High School in Sheffield trialled giving audio feedback to sixth-form pupils with very positive responses. Access to mobile devices was sufficiently good that this alternative form of feedback was readily accessible to pupils.

A pupil voice survey indicated that the majority of pupils scored the usefulness of the audio feedback higher than written feedback. The survey also indicated that the majority of pupils said they valued being able to listen back to feedback from their own teacher.

Teachers felt that pupils were better able to understand the audio feedback, teachers found it easier to give more audio feedback in a given time, and pupils were better able to pick up on verbal cues and tone of voice than nuances of written punctuation.

Description of the innovation

Teachers of Year 12 A-level mathematics and sociology trialled recording digital audio feedback and sharing this with pupils (for example via school email) instead of providing more traditional written feedback. Pupils accessed the feedback through mobile devices, and where this wasn't possible, through home computers.

Research questions

The specific primary question we addressed was:

“Is verbal (using an audio tool) feedback, delivered over a discrete topic, more effective than written feedback in improving test outcomes in sociology and mathematics A-level?”

And an important secondary question:

“Does providing verbal (using an audio tool) feedback to pupils rather than written feedback have a positive impact on teacher workload?”

If pupil outcomes are similar (ie, the answer to the primary question is inconclusive), there is still significant value for teacher retention in knowing whether or not providing audio feedback has a positive impact on teacher workload.

Method

Sample

TABLE 1: SCHOOLS INVOLVED IN THE EVALUATION

School identifier	Local Authority	School type	% FSM ever 6	% EAL	Participated in maths trial?	Participated in sociology trial?
A	Sheffield	Urban	28.1	20.4	Y	Y
B	Sheffield	Urban	20.2	27.9	Y	N
C	Sheffield	Urban	32.2	32.5	Y	Y
D	Warrington	Urban	10.6	2.6	N	N
E	Sheffield	Urban	11.3	8.4	Y	Y
F	Doncaster	Urban	22.7	12	N	Y
G	Cheshire East	Urban	16.2	3.1	Y	Y
H	Cheshire East	Urban	15.8	5.1	Y	Y

Although free school meal (FSM) and English as an additional language (EAL) data is included above for participating schools, this relates to key stages 3 and 4, whereas this study was carried out with Year 12 pupils who are in key stage 5 (KS5). We would anticipate the proportions of FSM and EAL pupils to be lower in KS5.

Participating schools were asked to provide two Year 12 maths classes and two Year 12 sociology classes to participate in the trial. Staffing changes in schools and pupil sixth-form option choices meant that there were not always two viable groups for each subject, so not all schools were able to provide two teaching groups for each subject.

The sample who participated in the evaluation had the following characteristics.

TABLE 2: MATHS TRIAL

	Intervention cohort	Control cohort
Number of pupils	92	89
Classes	4	5
Male	57%	64%
Female	43%	36%
Pupil Premium	1%	3%
Prior attainment (provided as average Attainment 8 score)	68.8	68.0

TABLE 3: SOCIOLOGY TRIAL

	Intervention cohort	Control cohort
Number of pupils	94	89
Classes	5	5
Male	17%	25%
Female	83%	75%
Pupil Premium	12%	12%
Prior attainment (provided as average Attainment 8 score)	52.6	52.2

Assignment to condition

Prior attainment data (average Attainment 8 score) was obtained for pupils from all participating groups prior to assignment to condition.

Assignment was carried out separately for the maths trial and the sociology trial. However, in each case the allocation process was the same. Entire teaching groups were randomly allocated to the intervention and control cohorts as below.

- Where a school provided two classes for a subject, one of the classes was randomly assigned to treatment with the other being assigned to control. This approach was taken in an attempt to eliminate as much as possible of the 'between-school variation' that we would expect to occur.
- Where a school only provided one class for a subject, the class was randomly assigned to either treatment or control.

The approach taken was that if randomisation did not produce similar prior attainment for intervention and control groups or similar cohort sizes, the process would be repeated.

In our evaluation randomisation was carried out twice for each study to produce intervention and control groups with similar GCSE attainment and cohort sizes.

Innovation

Subject specialists initially identified a maths topic and a sociology topic that would be taught by all exam boards at some point during Year 12. These topics would form the basis of the trial. The topics needed to be large enough to allow two to three pieces of work to receive either control feedback (written) or intervention feedback (recorded verbal).

- For mathematics the topic chosen was: *index laws, evaluating powers and differentiation*. This topic had the added benefit of appearing on all AS-level specifications. It was the view of the subject specialist that this is also a topic where pupils who are not able to respond effectively to feedback will make poor progress that will be reflected in any final summative assessment.
- For sociology the topic chosen was *research methods*. Variations between exam board case studies meant that it was sensible to restrict the trial to centres studying AQA sociology

Participating groups were given three doses of the intervention (either written or audio feedback) during the teaching of the identified topics. Teachers of classes in the written feedback condition provided typical written feedback on pupils' work. Teachers of classes in the audio feedback condition recorded spoken feedback as sound files; these files were emailed to individual pupils, who accessed the feedback on their mobile devices during the subsequent lesson.

Participating teachers (of both control and intervention cohorts) were instructed to provide 'appropriate feedback' on the three pieces of work that would be marked during the topic. This was to alleviate any concerns participating schools may have had about the compatibility of the trial with their individual assessment and marking policies. We also wanted the evaluation to reflect the variety of assessment and marking policies that schools have. The expectation was, therefore, that both intervention and control teachers would adhere to their own school assessment policies, but would deliver the marking and feedback in the two different ways prescribed.

We also required all participating schools to build in dedicated improvement and reflection time into lessons following the three pieces of feedback so that all pupils had an opportunity to respond to that feedback.

We also asked schools to ensure that where they had two groups in one subject (one control and one treated), they would endeavour to ensure that both groups otherwise followed as similar a scheme of work as possible for the topic concerned. This was to remove as much of the 'between-school variation' caused by different delivery approaches as possible.

We did not specify the length of time over which the innovation was to be delivered – largely to reflect the fact that different schools organise their Year 12 curriculums in different ways. In both mathematics and sociology, the subject specialists felt that the topics identified would be re-visited several times during the year and were substantial enough to ensure that there was opportunity for three pieces of work to be assessed. Again, by setting out to have one control and one intervention class in each subject in each school, we planned to remove as much of the effects of between-school variation as possible.

At the end of the teaching of the topic, all participating pupils (control and intervention) completed an assessment task. These were designed by our subject specialists and were in sealed envelopes until the assessment was carried out to avoid any teaching to the test. This will be referred to subsequently as the post-test.

Training was provided to all participating teachers at the outset. The items covered in training were:

- parameters for the trial
- administration of questionnaires during the trial
- running the final assessment
- dealing with consent and confidentiality
- dealing with the technical aspects of giving audio feedback
- what to do if there are any problems during the trial
- an opportunity for clarification on any potential issues
- time to explore the risks associated with running the trial – we felt that by better understanding the risks, participants were more likely to adhere strictly to the trial

protocols. These risks included demoralisation of control group, diffusion of treatment and managing the novelty effect for the intervention group.

The training was accompanied by the production of a comprehensive training manual. This included more detailed guidance around the scope of the topics selected by the subject specialists and suggestions of resources or activities that could be used as assessed tasks, although there was no expectation or requirement to use these particular activities.

Outcome measures

For maths pupils the pre-test measure used was their GCSE maths grade. For the sociology pupils the pre-test measure was their average Attainment 8 score. The use of a GCSE sociology grade as a pre-test measure would have been too severe a restriction on participation with such small numbers of pupils taking GCSE sociology.

The post-trial outcome was the assessed task for each subject. To address the fact that our own post-trial assessment was not as reliable or valid as the pre-test GCSE exam data, we carried out a comparative judgement ranking process using *No More Marking* software in order to obtain a post-trial score for each piece of work. This comparative judgement process was undertaken by all participating teachers from the trial and was carried out 'blind' in that all pupil work was anonymised prior to the comparative judgements being made. When combined with the fact that each piece of work was seen by multiple teachers in order to generate the post-test normalised score, the process was robust in terms of eliminating bias.

All 129 sociology post-tests were judged together. The binary decision data taken from the comparative judgements were statistically modelled by the *No More Marking* software to produce a score for each piece of work. The scores ranged from -4.99 to +4.49 with a mean of 0.27 and a standard deviation of 1.84.

Similarly all 151 mathematics post-tests were judged together with a score produced by the *No More Marking* software. The scores ranged from -6.84 to +4.24 with a mean of 0.03 and a standard deviation of 2.32.

We considered several outcome measures to evaluate the impact of the intervention on teacher workload as different school assessment policies have different expectations of teachers. We have made an assumption that workload related to teacher retention is as much about how teachers feel about the amount of work they have as it is about a clock-watching exercise. Therefore we asked participating teachers for their perception of the workload involved via the question:

What is your perception of the workload involved in providing feedback for this particular piece of work compared to the normal workload in providing feedback on work?

1=significantly less time than normal

2=less time than normal

3=same time as normal

4=more time than normal

5=significantly more time than normal

We anticipated that the control teachers would report no change, and the difference between these responses and those of the intervention teachers would help us answer the question

about the impact on workload. The teacher survey was conducted following each of the three assessed pieces of work.

Process evaluation

We also collected qualitative data from all pupils via Likert scale surveys following each of the three pieces of marked work. The questions specifically addressed the effectiveness of the feedback to the pupils and referenced known characteristics of good feedback as outlined in *What makes great teaching* (Coe, et al., 2014).

- I understand the feedback that I've been given.
- The feedback helped clear up any misunderstandings of what the question was asking me to do.
- The feedback helped by giving me specific things I could do to improve the work.
- The feedback told me the areas I could make improvements in, but left me to work out exactly what those improvements should be.
- Overall this feedback was useful.

The questionnaire also had an open question allowing for free-text responses.

Analyses

Primary question

Effect sizes of the treatment were calculated for each of the groups below:

- all sociology pupils
- male sociology pupils
- female sociology pupils
- disadvantaged sociology pupils
- all maths pupils
- male maths pupils
- female maths pupils
- disadvantaged maths pupils.

For each group that was analysed the pre-test scores and post-test scores were each scaled using the standard transformation $Z=(X-\mu)/\sigma$ so that the mean pre- and post-test scores for the group being analysed were scaled to 0, and the standard deviation to 1.

This allows the effect size to be calculated as the difference between the changes from pre- to post-test scores for each of the control and treated cohorts.

Secondary question

The quantitative Likert responses were plotted as average scores with 95% confidence intervals, with control and intervention cohorts plotted side by side. This allowed us to see whether there was a significant variation in the responses between control and intervention teachers in any of the cohorts. We plotted the responses following each of the three assessed pieces of work separately so that we could see whether teacher views changed over time as they became more familiar with the process. We also aggregated the teacher responses across the three assessed pieces of work to give us a larger dataset of responses.

This analysis was carried out separately for:

- all maths teachers
- all sociology teachers.

Analysis of the process

The quantitative Likert responses from pupils were plotted as average scores with 95% confidence intervals with control and intervention cohorts plotted side by side. This allowed us to explore any significant variation in the responses between control and intervention pupils in any of the cohorts. We plotted the responses following each of the three assessed pieces of work separately so that we could see whether pupil views changed over time as they became more familiar with the process. We also aggregated the pupil responses across the three assessed pieces of work to give us a larger dataset of responses.

This analysis was carried out separately for:

- all maths pupils
- all sociology pupils.

Although our project paperwork asked pupils to identify whether they were responding to questions in lesson one, two or three on the proformas, we had a small number of returns which did not include the lesson number. These Likert responses were included within the aggregated pupil responses, but were obviously not used in the analysis for individual pieces of work.

Pupil responses to the open question were grouped into themes.

Cost

The cost of the project is broken down below.

The total project cost was £25,775. This was lower than the original granted figure due to attrition of teaching groups. This meant that we paid participating schools less for attendance at training and also for attendance at the final evaluation.

This can be broken down as £11,950 for costs associated with:

- project management
- employment of subject specialists to identify appropriate topics and create an appropriate final assessment
- creation of all training materials
- delivery of training
- project evaluation including data manager time
- administration.

And as £13,825 paid to participating schools to cover:

- staff time to attend training
- staff time to attend final comparative judgement moderation
- travel
- administration
- internal technical support to enable recording of audio files.

The cost of delivering the intervention (ie, project costs minus costs related to the evaluation) is calculated as £6,925.

This can be broken down as £1,500 for the costs of providing training and £5,425 as release and travel time for participating teachers along with technical support at the participating schools (based on 21 teachers from seven schools). We have included the costs of training teachers who did not participate in the evaluation due to changes to schools' A-level blocks and timetables between the training for the trial and the trial itself.

Therefore the cost 'per pupil' can be calculated as £25.

Once a teacher has been trained in the intervention, there is no need for additional intervention/resources/training and therefore the costs per pupil would fall significantly over time.

Results

Outcome findings for primary research question

Effect sizes are calculated in Tables 4 and 5.

The number of data points is lower than the total number of pupils in the trial due to missing data – pupils not present for post-test due to changing A-level options or absences.

TABLE 4: SOCIOLOGY

Cohort	Control	Treated	Effect size
Entire study	-0.08 (N=60)	+0.07 (N=69)	+0.15
Boys	-0.10 (N=14)	+0.12 (N=12)	+0.22
Girls	-0.12 (N=46)	+0.10 (N=57)	+0.22
Disadvantaged	+0.28 (N=8)	-0.36 (N=9)	-0.64

TABLE 5: MATHS

Cohort	Control	Treated	Effect size
Entire study	-0.10 (N=63)	+0.08 (N=78)	+0.18
Boys	-0.08 (N=43)	+0.07 (N=47)	+0.15
Girls	-0.09 (N=20)	+0.08 (N=31)	+0.17
Disadvantaged	-	-0.15 (N=1)	-

With the exception of disadvantaged pupils the data suggests that there was a positive effect size for all cohorts undergoing the treatment. Further investigation would be useful to establish why disadvantaged pupils had such a negative effect size.

Outcome findings for secondary research question

What is your perception of the workload involved in providing feedback for this particular piece of work compared to the normal workload in providing feedback on work?

1=significantly less work than normal

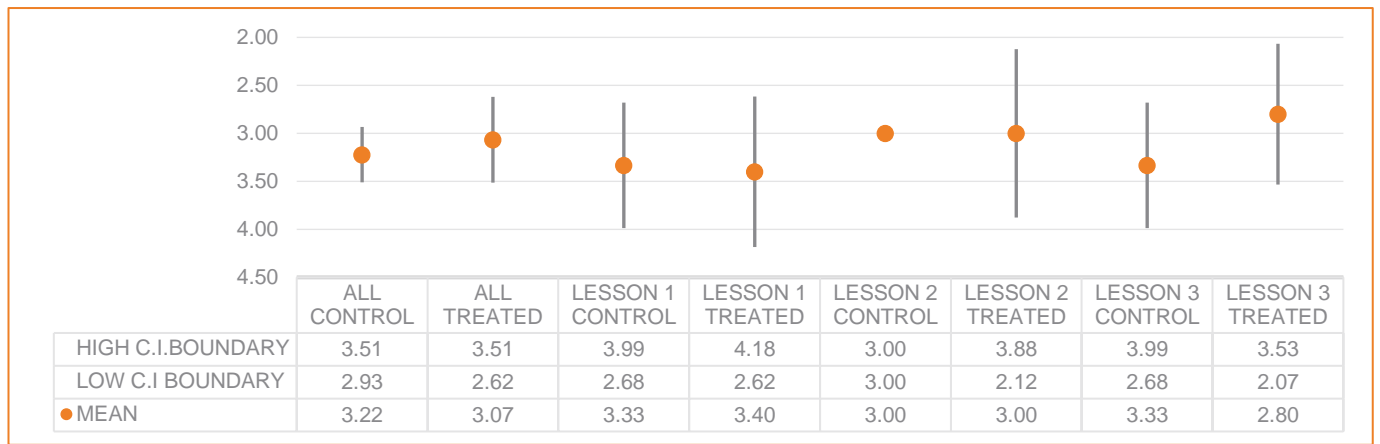
2=less work than normal

3=about the same work as normal

4=more work than normal

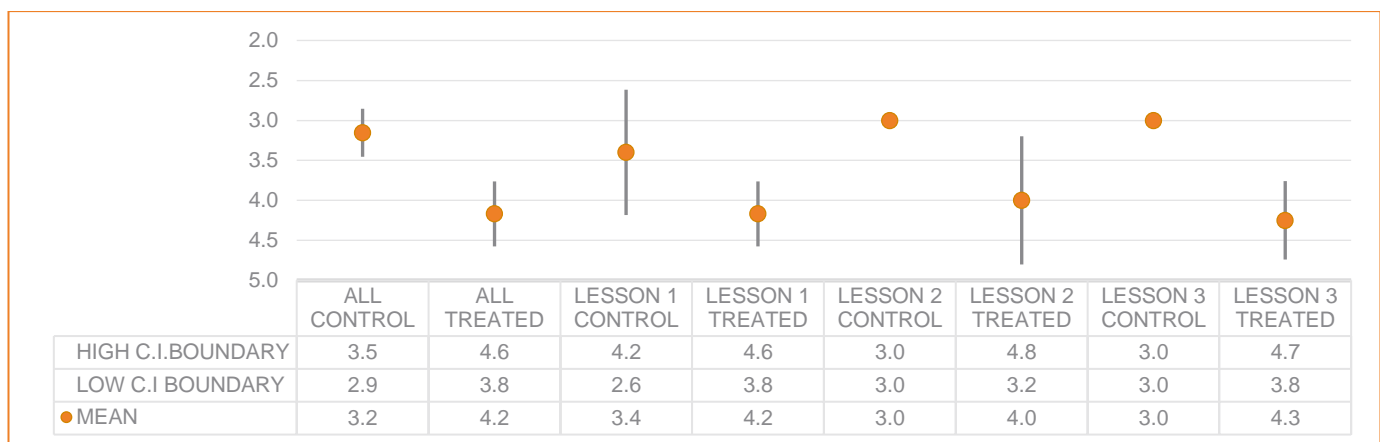
5=significantly more work than normal

CHART 1: SOCIOLOGY



Control (N=5), Treated (N=5).

CHART 2: MATHS



Control (N=5), Treated (N=4)

The numbers of teachers involved in these questionnaires is small and therefore calculations around significance need to be treated with extreme caution.

The first observation is that while we would have expected control responses to have all been '3', some teachers have responded that this involved more work than normal. This could be due to conflating administrative work involved with adherence to trial protocols with the work involved with giving feedback.

Sociology teachers involved with the trial seem not to have considered the process of giving audio feedback to take any more time than giving normal feedback. What is interesting to note is that by the end of the trial the teachers of the treated groups seem to become increasingly favourable towards audio feedback when considering the time taken. This could be because they have got over initial technical issues and have become better and more efficient at doing so (teacher comments support this theory). A trial over a longer period would establish whether this trend continues. A one-tailed t-test on the aggregate responses for sociology gives a p-value of 0.28.

The responses from maths teachers are statistically significant with teachers perceiving that audio feedback takes longer than normal feedback. Additionally, there is no trend over time with

similar discrepancies in responses between teachers of control and treated groups over the duration of the trial. A one-tailed t-test on the aggregate responses for maths gives a p-value <0.001 supporting the significance of this finding.

The difference in responses from the two subject areas could be to do with the fact that maths can often be marked quickly, with brief annotations often providing the necessary feedback and identifying common errors or misconceptions, whereas essay-based subjects such as sociology may often require more personalised feedback that takes longer to give, and therefore there is more scope to give feedback in alternative forms without taking any longer to do so.

Useful insight was also obtained through discussion with teachers at the moderation of post-trial assessments.

Comments were made on the following themes:

- Technical issues – one teacher commented that they had discovered a way of inserting pauses into the recordings that allowed the teacher unpressured thinking time and avoided long gaps in the feedback.
- Experiences of improving the quality of audio feedback they were providing. One teacher described how after the first run-through they decided to identify in advance key aspects of the work that they would focus on to result in more targeted comments (this would be a feature of good feedback anyway and suggests that the process of engaging with audio feedback has helped the participating teacher reflect on their own practice). Another teacher described how they became better over the trial at giving specific comments rather than long descriptions.
- Experimenting with different practice – one teacher tried withholding a grade until the end of the comment or after the feedback was received and found that the pupils engaged with the audio feedback more as a result. This is perhaps unsurprising, but again shows the introduction of a new means of giving feedback causing a teacher to reflect on their practice.
- Becoming more comfortable and fluent in their comments – again unsurprising, but teachers commented that they initially felt awkward providing the audio feedback but that they became much less so as the trial proceeded.

Any advice to teachers wanting to use audio feedback would include suggesting that they reflect on the above comments before using this as an intervention tool.

Process evaluation findings

Analysis of the pupil Likert responses are given below.

Question 1:

I understand the feedback I've been given

1 = strongly agree

2 = agree

3 = neither agree nor disagree

4 = disagree

5 = strongly disagree

CHART 3: SOCIOLOGY

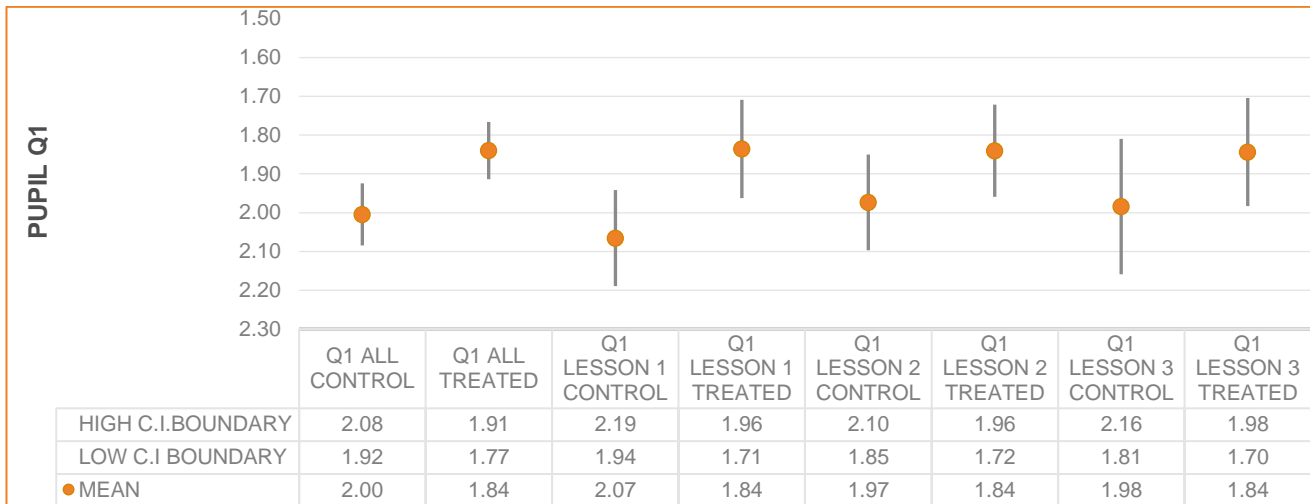
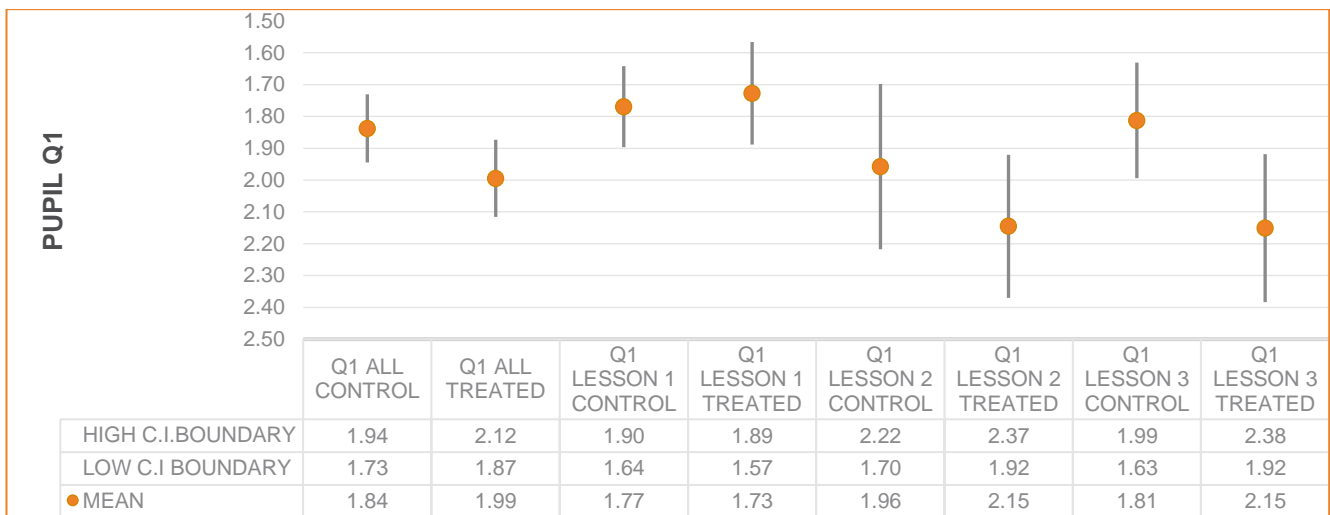


CHART 4: MATHS



Results depend on the subject. The aggregate results for sociology pupils suggest that there is a statistically significant difference in responses from the pupils, with the pupils better understanding the feedback from the treatment. For maths, the results again have a statistically significant difference, this time showing pupils understanding the feedback from the treatment less well than conventional feedback.

One-tailed t-tests performed on the aggregate responses for sociology and maths provide p-values of 0.002 and 0.029 respectively which support the above interpretation.

Question 2

The feedback helped clear up any misunderstandings of what the question was asking me to do.

1 = strongly agree

2 = agree

3 = neither agree nor disagree

4 = disagree

5 = strongly disagree

CHART 5: SOCIOLOGY

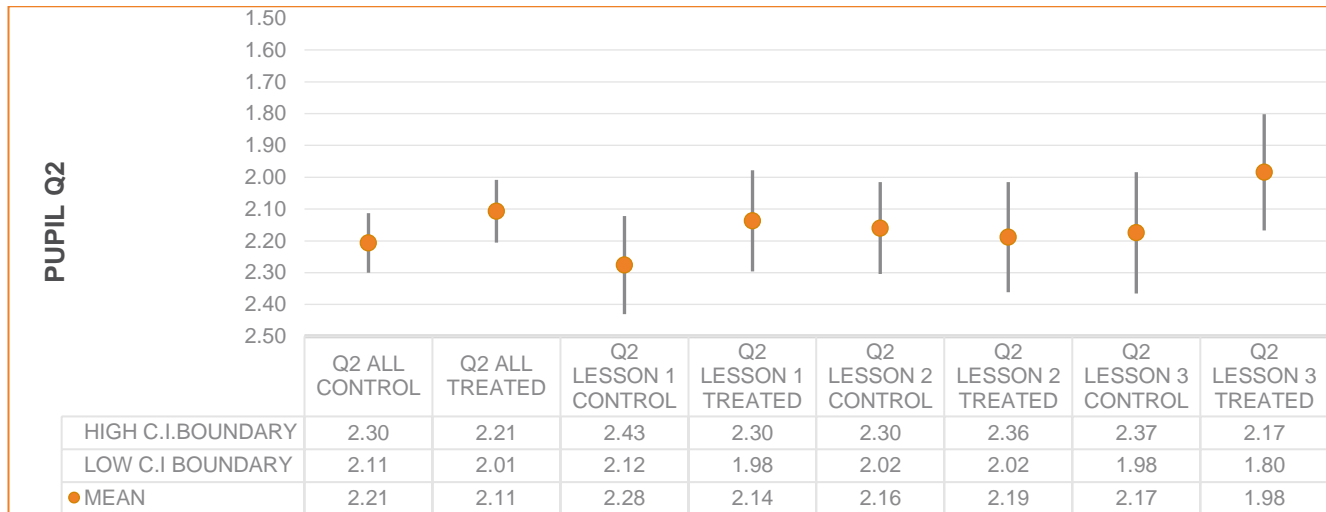
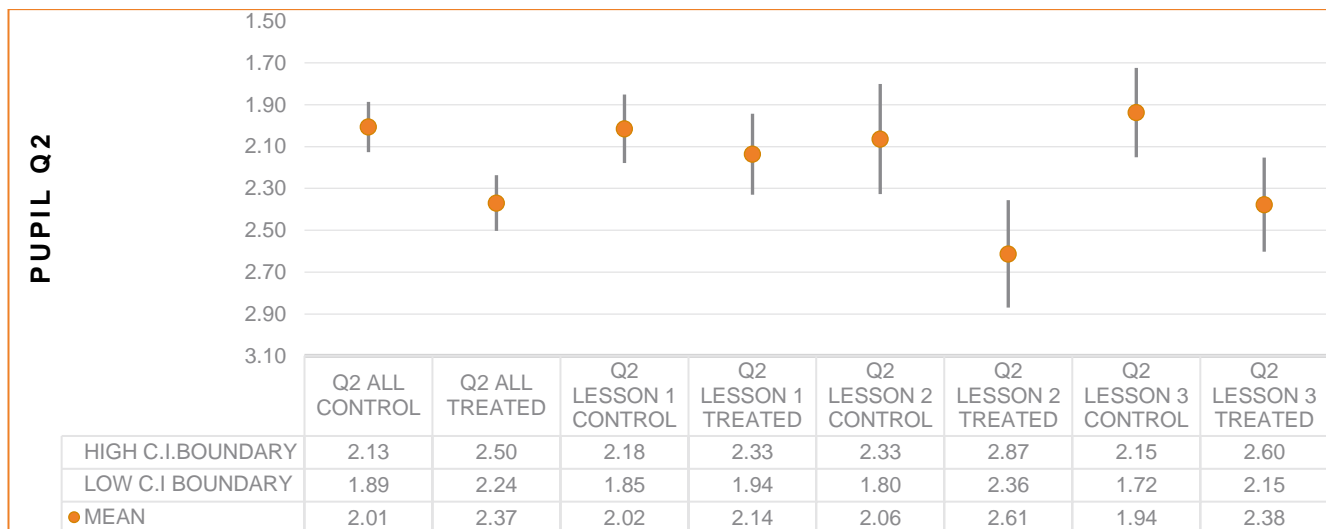


CHART 6: MATHS



The results indicate that sociology pupils feel that the audio feedback was slightly better than conventional feedback at helping them better understand what the task was asking to do. For maths however, there was a statistically significant perception the other way. This could be to do with the nature of the tasks for sociology and maths. Maths tasks invariably require pupils asking a specific question, with a clear 'answer' requested, whereas a sociology task may be more open-ended in nature.

One-tailed t-tests performed on the aggregate of responses for sociology and maths give p-values of 0.076 and <0.001 which supports the above interpretation.

Question 3:

The feedback helped by giving me specific things I could do to improve my work

- 1 = strongly agree
- 2 = agree
- 3 = neither agree nor disagree
- 4 = disagree
- 5 = strongly disagree

CHART 7: SOCIOLOGY

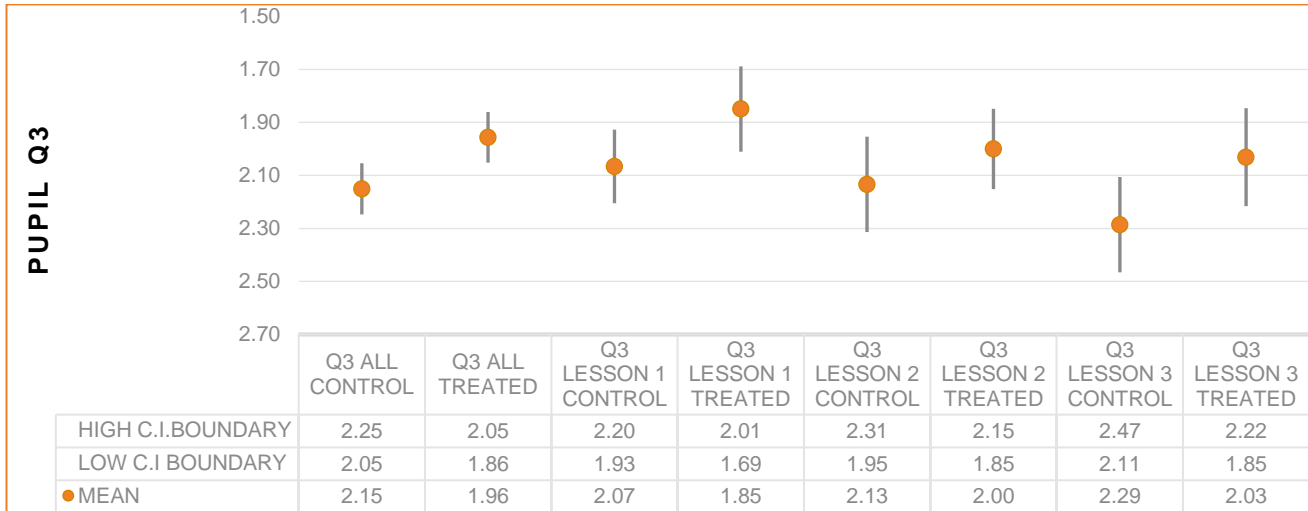
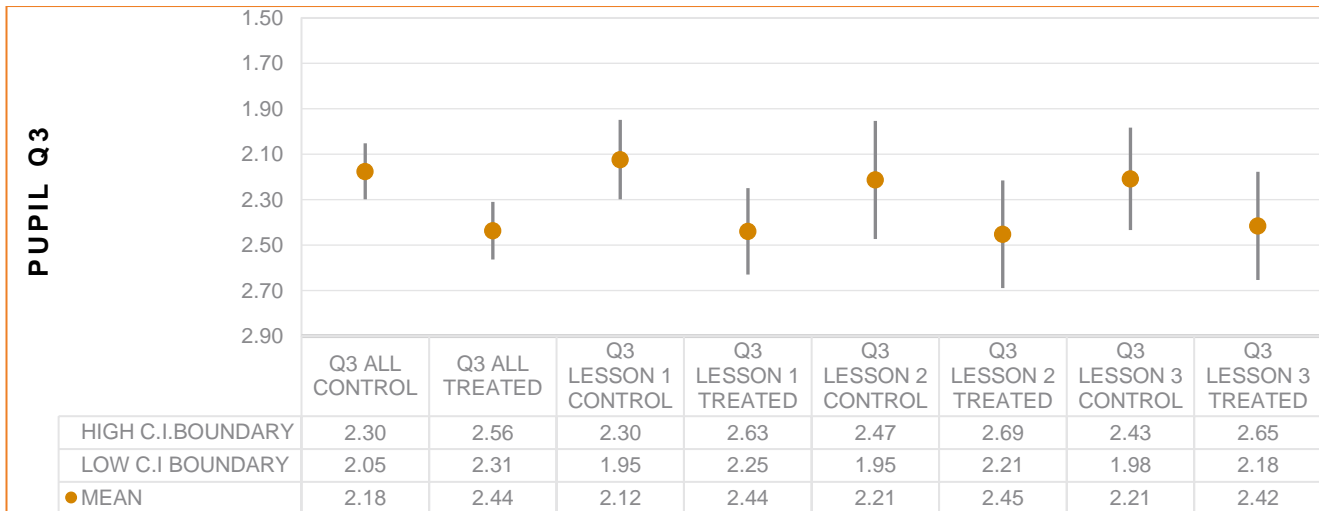


CHART 8: MATHS



For both subjects the results are consistent, both following each task and when considering the aggregation of feedback. In both cases the aggregated results are also statistically significant. In the case of sociology, there is a statistically significant preference for audio feedback, whereas in maths the result is equally strong but in favour of conventional feedback. This could be to do with the fact that within a maths task there are usually a relatively small number of routes to take through a problem with written feedback often able to identify a 'common mistake'. There would usually be a greater degree of variation in more open-ended written tasks associated with sociology.

One-tailed t-tests performed on the aggregate responses for sociology and Maths give p-values of 0.003 and 0.002 respectively to support the above interpretation.

Question 4

The feedback told me the area I could make improvements in but left me to work out exactly what improvements to make

- 1 = strongly agree
- 2 = agree
- 3 = neither agree nor disagree
- 4 = disagree
- 5 = strongly disagree

CHART 9: SOCIOLOGY

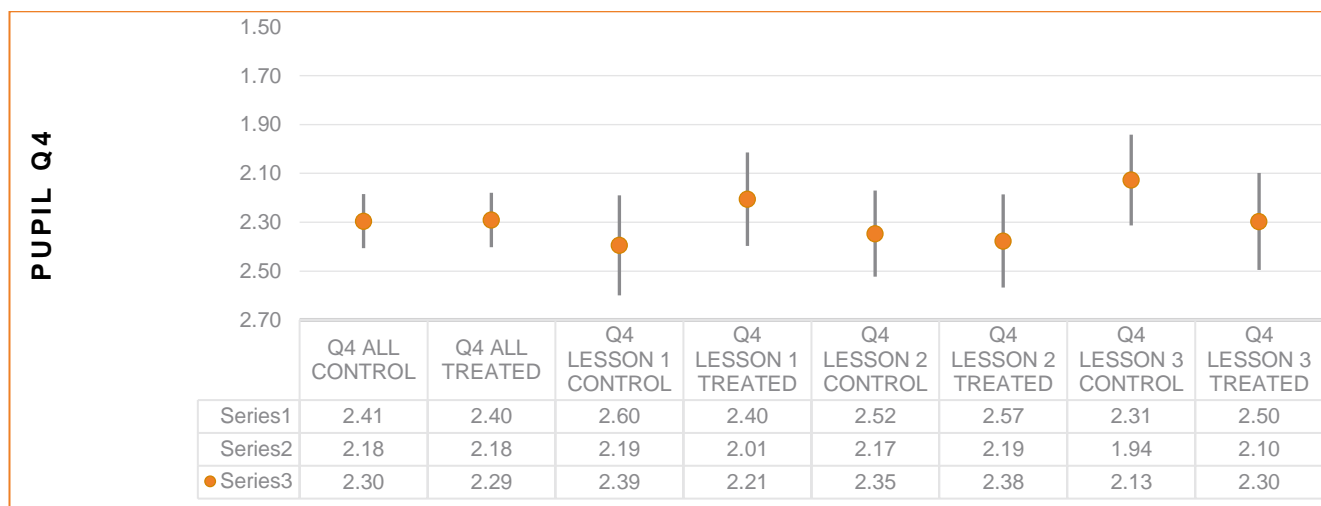
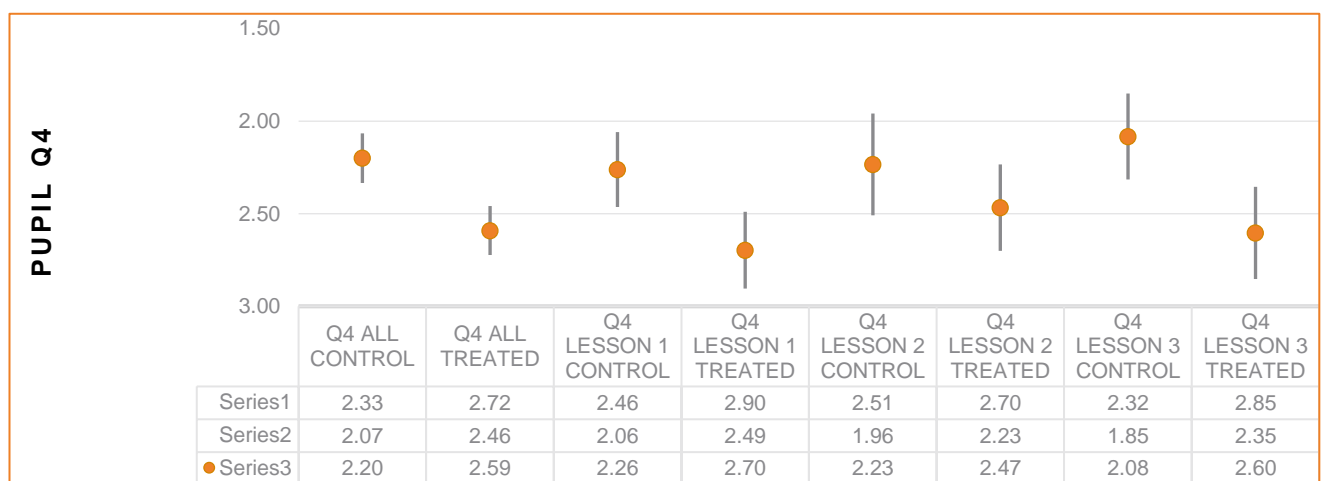


CHART 10: MATHS



The results for maths are again statistically significant and favouring conventional feedback. The pupil responses from sociology show no variation in pupil responses between control and treatment.

One-tailed t-tests performed on the aggregation of pupil responses for sociology and maths give p-values of 0.48 and <0.001 respectively to support this interpretation.

Question 5

Overall this feedback was useful

1 = strongly agree

2 = agree

3 = neither agree nor disagree

4 = disagree

5 = strongly disagree

CHART 11: SOCIOLOGY

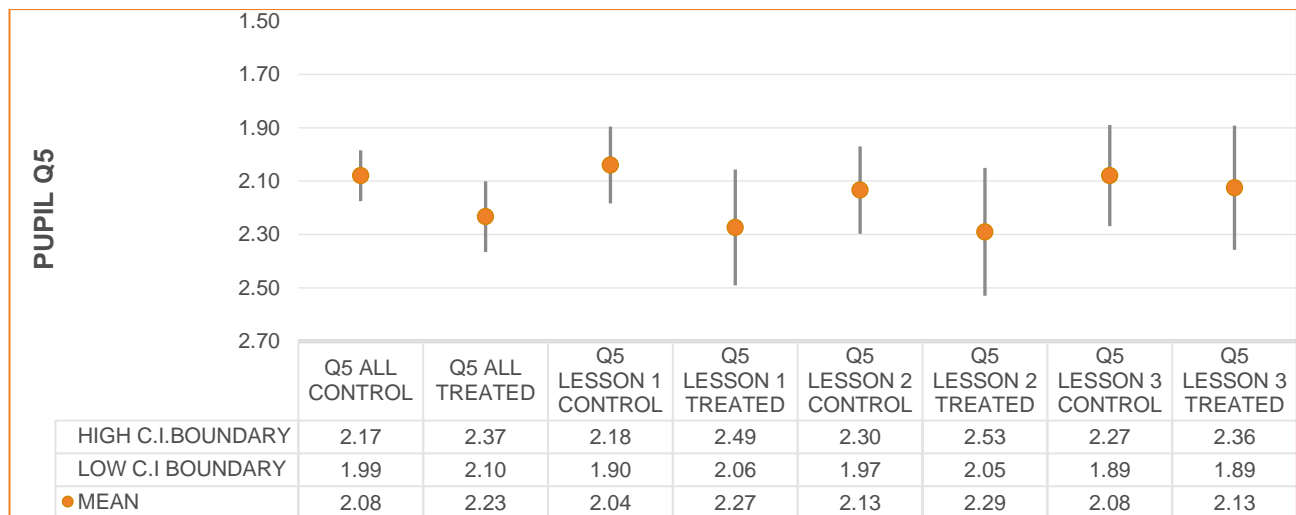
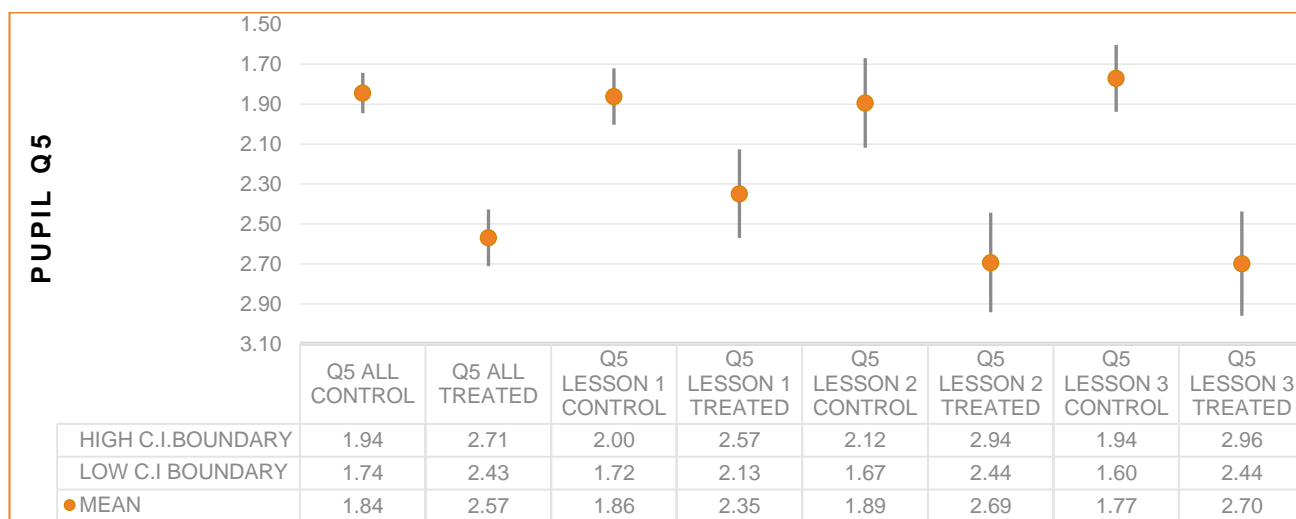


CHART 12 MATHS



For maths the results are not surprising given pupil responses to earlier questions. The pupils have consistently found conventional feedback more useful. The data suggests that maths pupils' perceptions of the respective utility of the feedback diverged as the study progressed.

The pupil response for sociology for this final question are also significant and are surprising given the responses to the earlier questions.

One-tailed t-tests performed on the aggregate of responses for sociology and maths give p-values of 0.03 and <0.001 which support the above interpretation.

Free text responses

Comments from pupils can be grouped into themes:

- Visibility of the feedback for future reference. Pupils commented that the audio nature of the feedback meant that it wasn't easily accessible to them when they looked back through their work. This would be something to consider if implementing audio feedback as a strategy – one suggestion would be to ask pupils to listen to the audio feedback and then annotate their own work.
- Taking longer to access the feedback.
- Technical issues. A number of pupils commented that the audio files wouldn't work on their phones and it therefore took longer to access the feedback.
- Difficulties 'visualising' the feedback – this was specific to maths.

Discussion/Conclusion

The answer to our primary research question is that, in this study, the use of verbal feedback (via an audio tool), delivered over a discrete topic, was more effective than written feedback in improving test outcomes in sociology and mathematics A-level

The overall effect size for sociology was +0.15 and for maths +0.18. Effect sizes were similar for both boys and girls.

The effect size for disadvantaged pupils in sociology was -0.64. There was insufficient data to draw conclusions for effect size for disadvantaged pupils in maths. The reason for this negative effect size could be due to the various challenges pupils described in accessing audio feedback easily. This may typically present more challenges to disadvantaged pupils than other pupils and strategies to counter this would need to be explored. With the caveat that the number of disadvantaged pupils involved in the evaluation was small (and therefore these findings should be treated with caution), teachers who are considering this approach in a class with a larger number of disadvantaged pupils should take care over implementation and evaluation.

The study revealed interesting survey responses from pupils with a strong split in responses from sociology and maths pupils.

At a statistically significant level the sociology pupils reported understanding the audio feedback better, found the audio feedback more useful at providing specific direction around improvement, but surprisingly (in light of their other responses) stated that overall they found the audio feedback less useful than conventional written feedback.

At a statistically significant level the maths pupils reported understanding conventional feedback better, found conventional written feedback more helpful at clarifying what the task was asking, found conventional written feedback more helpful in providing specific direction around improvement, found conventional written feedback more helpful in identifying general areas to focus on for improvement, and overall found conventional written feedback more useful than audio feedback.

The answer to our secondary research question is that providing verbal feedback (using an audio tool) to pupils rather than written feedback may have a positive impact on perceived teacher workload for sociology teachers, and has a statistically negative impact on perceived teacher workload for maths teachers.

In conclusion, in this study the use of an audio feedback tool in maths is not an improvement on conventional written feedback. Although the effect size for maths outcomes was positive, there was a statistically significant negative impact on perceived staff workload and pupil voice was clear that that they preferred conventional written feedback.

In sociology, this study suggests that the use of an audio feedback tool is an improvement on conventional written feedback. The effect size for sociology outcomes was positive, teachers reported that there may be an improvement in workload and pupils found the feedback useful. There are some aspects of delivery that would need addressing prior to employing this method of feedback in sociology classrooms – namely adapting teaching strategies to address both feedback from pupils about the accessibility of feedback and associated technical issues. Similarly further work would be needed to ensure that the needs of disadvantaged pupils are fully met.

It would be reasonable to hypothesise that the approach of giving audio feedback to pupils may be equally effective in A-level subjects that are similar to A-level sociology in terms of the type of conventional feedback that is usually given. For example, there are a number of subjects at key stage 5 where pupils need to provide longer prose answers to questions. All may be suitable for evaluation of the use of audio feedback and further evaluation would be valuable to establish both the effectiveness of audio feedback and staff and pupil attitudes in these similar subject areas.

It would also be valuable to further explore the reason why there was a substantial negative effect size for disadvantaged pupils and what could be done to mitigate the potential negative effect for this group.

Within our own school these results will be shared with heads of department who will be encouraged to trial similar approaches. We will also share the findings with those schools involved in the trial who we anticipate will disseminate these results similarly. We also look forward to sharing our findings with those networks that we are actively engaged with.

References

Coe R, Aloisi C, Higgins C, & Elliot Major, L, 2014. What makes great teaching, England: Sutton Trust.

DfE, 2015. Government response to the workload challenge, London: Department for Education.

Elliott V et al., 2016. A marked improvement, England: EEF, Oxford University.

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



Contact us

+44 (0)1904 328166 info@the-iee.org.uk
Berrick Saul Building, University of York, York YO10 5DD
Twitter: [@IEE_York](https://twitter.com/IEE_York) the-iee.org.uk/

© Institute for Effective Education, 2019

The Institute for Effective Education (IEE) is an independent charity working to improve education for all children by promoting the use of evidence in education policy and practice.

In collaboration with the Education Endowment Foundation (EEF) we support a national Research Schools Network and have developed resources aimed at people on the front line of education.

The Institute for Effective Education is a charity registered in England, charity number 1168744

Institute for
Effective Education
Empowering educators with evidence

