Visible Classroom: re-grant

Evaluation Report

February 2021

Michael Sanders, Aisling Ní Chonaire, Francesca Tamma, Bibi Groot, Daniel Carr, Jessica Heal, Juliane Wiese, Sarah Breathnach, Lauren Crouch, Matthew Barnard, Hazel Wright, and James Lawrence.

The Education Endowment Foundation (EEF) is an independent grant-making charity dedicated to breaking the link between family income and educational achievement, ensuring that children from all backgrounds can fulfil their potential and make the most of their talents.

The EEF aims to raise the attainment of children facing disadvantage by:

- identifying promising educational innovations that address the needs of disadvantaged children in primary and secondary schools in England;
- evaluating these innovations to extend and secure the evidence on what works and can be made to work at scale; and
- encouraging schools, government, charities, and others to apply evidence and adopt innovations found to be effective.

The EEF was established in 2011 by the Sutton Trust as lead charity in partnership with Impetus (formerly Impetus Trust) and received a founding £125m grant from the Department for Education.
Together, the EEF and Sutton Trust are the government-designated What Works Centre for improving education outcomes for school-aged children.

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

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

0207 802 1653

jonathan.kay@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Contents

About the evaluator

The project was independently evaluated by a team from the Behavioural Insights Team: Michael Sanders, Aisling Ní Chonaire, Francesca Tamma, Bibi Groot, Daniel Carr, Jessica Heal, Juliane Wiese, Sarah Breathnach, Lauren Crouch, Matthew Barnard, Hazel Wright, and James Lawrence.

The lead evaluator was Michael Sanders.

Contact details:

Michael Sanders
The Behavioural Insights Team
4 Matthew Parker Street
London
SW1H 9NP.

Executive summary

The project

The Visible Classroom intervention supports teachers' professional practice development by providing personalised feedback and mentorship on teaching practice with the aim of improving student mathematics and reading outcomes in Years 5 and 6 (ages 9 to 11). Teachers received a word-for-word transcript of each of their lessons that they recorded and submitted to the programme platform and were able to access a data dashboard summarising the lesson statistics (for example, teacher talk time, speed of speaking, and the number of teacher and student questions asked). In addition, they received a feedback report that assessed their lessons and offered considerations for improvement designed to encourage critical reflection and improvements in classroom practice. Other resources for teachers included links to further information and collaborative professional dialogue sessions with a nominated teaching mentor.

Teachers from intervention schools took part in a one-day in-person training session in which they were introduced to the pedagogical model, the technological platform, and the ways in which the feedback could be applied to their teaching practice and self-selected teaching goal. In addition, teachers received a copy of a training manual that covered the materials that were discussed in the training. Teaching mentors, who served to help teachers reflect on their teaching feedback throughout the academic year, were trained through an hour-long webinar to learn about effective mentorship processes.

In total, teachers of 7,230 students representing 86 schools took part in the evaluation. This clustered randomised controlled trial ran from April 2016 to July 2018 and compared treatment schools to a 'business as usual' control. An implementation and process evaluation assessed fidelity and other aspects of the programme using pre- and post-intervention surveys amongst teachers who participated from 2016 to 2018 and were willing to respond in the form of online questionnaires, semi-structured interviews, and observed training sessions.

Key conclusions	
1.	Pupils in schools that received Visible Learning made, on average, one month less progress in combined KS2 maths and reading compared to children in the control schools. This is our best estimate, which has a high security rating. Variability around this best estimate means that possible impacts of the study include larger negative effects and negligible positive effects.
2.	A negative impact of two months reduction in learning was estimated for KS2 reading outcomes for students in Year 5. The variability around this best estimate means that, while the possible results are consistently negative, the magnitude may be larger or smaller than a two month reduction.
3.	Among children eligible for free school meals (FSM), those in the treatment schools made no additional progress in the KS2 maths and reading outcome measure compared to those in the control groups. These results may have lower security than the overall findings because of the smaller number of pupils. Variability around the estimate indicates that possible effects range between substantial positive and negative impacts.
4.	The IPE found that iterative feedback was especially well received in schools with an open working culture that encouraged feedback compared to schools where feedback was emphasised less.
5.	There were a number of implementation barriers that may have hindered the success of the intervention. Examples included teachers not having enough time to properly engage with the feedback they were given and, in some cases, feeling self-conscious about lessons being recorded.

EEF security rating

These findings have a high security rating. This was an efficacy trial, which tested whether the intervention worked under developer-led conditions. The trial was a well-designed two-armed randomised controlled trial. The trial

was well-powered and relatively few pupils (5%) who started the trial were not included in the final analysis. The pupils in each group had similar prior attainment at the beginning of the trial. The trial lost one padlock (see Table 1) due to a lack of information on practice within the schools in the comparison group.

Additional findings

The intervention was found to have no positive effect on the combined KS2 maths and reading outcomes for Years 5 and 6, nor on the KS2 maths outcomes for Years 5 and 6 or the KS2 reading outcomes for Year 6. The intervention was found to have a negative effect on KS2 reading outcomes for Year 5 resulting in a loss of approximately two months' progress. This result is statistically significant at the 5% level with the possible impacts ranging between effect sizes of -0.01 (0 months) and -0.25 (-3 months).

Teachers experienced some self-consciousness and discomfort about recording their lessons. The impact of this was that some teachers reported delaying recording or uploading their lessons. Teachers expressed mixed feelings about the need for in-person training: some felt it was essential while others wondered whether there were more cost-effective methods of disseminating the same information. Training seemed to be a useful guide on how to use the programme app but other aspects of participation, such as understanding how long the recordings needed to be, what to do when the class is interrupted, and the exact role of the mentor were still unclear to participants after the training. Teachers also reported initial issues recording lessons via their smartphones.

A number of barriers to ongoing participation emerged. Some teachers found the strict deadlines difficult to meet. Teachers mentioned that work and home life commitments affected their ability to regularly record lessons. Teachers also reported confusion as to whether or not there were strict criteria regarding which lessons they should record. Many teachers interviewed valued the objective style of the feedback and felt that it was evidence-based and data-driven. The fact that the feedback was seen as objective helped them accept it as a more accurate portrayal of their performance and therefore easier to take on board. Timing of feedback delivery proved important with immediate feedback being more beneficial. Teachers found it challenging to apply the feedback to different contexts, for example, applying feedback given about an English lesson to a maths lesson later in the term. While the process evaluation did examine factors around implementation, it did not seek to measure behaviour change or other teacher-based outcomes.

Cost

The average cost per pupil over three years for Visible Learning is £23. Most of this expense relates to set-up costs in the first year, which means that the cost per pupil decreases over the years. The costs relate to training, staff cover, and, in some cases, buying equipment.

Impact

Table 1: Summary of impact on primary outcome

Outcome/ Group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	No. of pupils	P value	EEF cost rating
KS2 maths and reading, Year 5	-0.09 (-0.22, 0.04)	-1	4	3,337	0.191	£££££
KS2 maths and reading, Year 6	-0.05 (-0.19, 0.09)	-1	4	3,597	0.549	£££££
KS2 maths and reading, Year 5—FSM pupils	-0.03 (-0.23, 0.17)	0	N/A	1,232	0.772	£££££
KS2 maths and reading, Year 6—FSM pupils	0.03 (-0.13, 0.20)	0	N/A	1,344	0.708	£££££

Introduction

This report presents the findings of the evaluation of Visible Classroom, an approach to supporting teachers through personalised feedback. This first chapter provides background to the development of the intervention and a detailed description of how it is intended to be implemented. The chapter then goes on to set out the evaluation objectives and summarises the trial's ethical review and data protection provisions before finally providing details of the development and intervention teams.

Background evidence

Quality of teaching has been shown to be a key determinant of learning outcomes and is second only to individual student effort (Rowe, 2003). Indeed, an estimated 30% of the variance in student achievement is explained by teachers alone (Hattie, 2003). Within this context, there is good evidence that timely, appropriate feedback can improve teaching and that 'effective instruction cannot take place without proper feedback from student to teacher on the effectiveness of the instruction' (Arnold, 2011). The wider literature on feedback suggests that it encourages individuals to reflect on their learning, helps them to contextualise their progress, and clarifies areas for improvement. Importantly, feedback provides the opportunity for self-reflection and self-assessment of skills and capabilities (Alberts, Martijn, and De Vries, 2011). Further research takes advantage of teacher observation to identify what type of feedback is of most value in the education space. One study found that certain qualities of teacher interactions predicted higher levels of student attainment, for example, fostering a positive emotional climate, attentiveness to individual needs, and engaging with a variety of teaching models were particularly important (Allen et al., 2013). In this study, teacher observations were used to identify particular needs in educational settings, however, the use of observations (and feedback) as the intervention itself is a novel extension of the existing literature in this space.

In 2009, John Hattie, the educational researcher who helped to develop the Visible Classrooms intervention, published a synthesis of over 800 meta-analyses relating to educational achievement. The findings of this synthesis highlighted that activities associated with 'visible teaching and learning' are consistently associated with improved attainment among students (Hattie, 2009). According to Hattie, for visible teaching and learning to occur, an explicit and appropriately challenging goal first needs to be identified by student and teacher. Once this goal is set, both parties should engage in deliberate activities to practice skills that will result in mastery of that goal. Visible teaching and learning will occur when feedback is sought and given and when both student and teacher are engaged and passionate about the learning process (Hattie, 2009).

Drawing on these key findings, John Hattie, Janet Clinton, and their team at the University of Melbourne developed Visible Classroom. Visible Classroom is an education intervention that aims to improve student attainment in reading and maths by supporting the development of teachers' professional practice. The intervention uses real-time captioning technology developed in collaboration with Ai-Media UK to create transcripts of lessons delivered by teachers. These transcripts are analysed by a team of specialists at the University of Melbourne who provide personalised feedback within a few weeks to teachers on key features of their lessons such as thinking time provided to students, questioning style used, and goals identified during the teacher training. This feedback, along with the full lesson transcripts, is provided via a personalised teacher dashboard on the programme app. Teachers can also share the transcripts with their class, enabling students to review learning instructions, goals, and to provide feedback on each lesson.

In 2014, a pilot evaluation of the Visible Classroom intervention was conducted with ten primary schools in London and the West Midlands. This pilot study sought to assess the feasibility of the Visible Classroom technology and gather evidence on the potential effectiveness of the overall approach (Skipp and Tanner, 2015). Using a combination of semi-structured interviews with staff and classroom observations, Visible Classroom was deemed to be a feasible intervention with potential to improve teacher practice and eventually student attainment. The evidence from the pilot indicated that although very few teachers spent time reviewing the entire transcript of their lessons, teachers used the dashboard to view transcript statistics regularly and felt that this was the aspect of the approach most likely to change their behaviour and lead to improvements in pupil attainment. There was,

however, less evidence that making live transcripts of lessons available to pupils would be beneficial. Teachers also had mixed views about which pupils were likely to have benefited most from the live transcripts.

The theory of change, as described in the pilot evaluation report, outlines a cyclical, four-step process (Skipp and Tanner, 2015). The process begins with improved visibility of teaching and learning, that is, appropriate goal setting, deliberate practice of the goal, student-teacher feedback loop, and engagement in the learning process. Improved visibility of teaching and learning is followed by improved feedback to teachers, change in teacher practice, and finally improved student learning outcomes. The fundamental idea is to prompt 'changes in teacher practice and mindset, such that teachers become evaluators of their own teaching … leading to increased attainment, particularly amongst disadvantaged students' (Skipp and Tanner, 2015).

The objective of the current evaluation was to assess the efficacy of the Visible Classroom intervention under ideal conditions (for example, when being delivered to teachers, as intended, by the intervention's original developer). Part of the rationale for conducting this evaluation was also to assess whether Visible Classroom could be scaled up and introduced across a wide range of schools in the U.K. To date, no other work has explored the influence of the Visible Classroom intervention on student outcomes, although there are schools in Australia, the U.S.A., the U.K., and Germany that have one or more teachers participating in Visible Classroom.

Intervention

As indicated above, the Visible Classroom programme is designed to support teachers using personalised feedback on their lessons combined with mentorship aimed at supporting them to find ways to apply the feedback to their practice. The description of the intervention below follows the Template for Intervention Description and Replication (TIDiER) checklist (Hoffman, 2014).

1. **Brief name**: The Visible Classroom (VC).

2. **Why (rationale/theory)**: The intervention aimed to improve student achievement outcomes by supporting teachers' professional practice development. The proposed mechanism by which this improvement occurred was via the provision of personalised feedback on teaching. This feedback was designed to encourage teachers to reflect critically on their teaching and develop their classroom practice in accordance with their teaching goals, which would be expected to have flow-on effects for their pupils' learning and attainment in reading and maths.

3. **What (materials)**: Mentor teachers, non-participating teachers in intervention schools, were nominated by participating teachers such that there were approximately equal numbers of teachers and mentors in a school. Mentors were encouraged to participate in an online webinar. In this interactive session, mentors were introduced to the VC intervention as well as the mentor's role and expectations. Mentors also received guidance on developing relationships with teachers, supporting teacher practice and development, and practical tips and suggested protocol for doing so.

   Teachers took part in a one-day, in-person training session in which they were introduced to the pedagogical model, the technological platform, and the ways in which the feedback could be applied to their teaching practice and self-selected teaching goal. Professor John Hattie, one of the developers of the VC intervention, presented on the project background and research and explained the teachers' role in participating in the VC intervention. Teachers also received guidance on how to interpret feedback reports, working with mentors, and completing pre- and post-intervention surveys. In addition, teachers received a copy of a training manual—which covers the materials that were discussed in the training—and links to online training materials, videos, and guides. Sample feedback reports were provided along with a document for understanding these metrics.

   Teachers used a smartphone or tablet app to record lessons and uploaded them to their own personal profile on the VC platform. Within a few weeks of uploading, teachers received a transcript of their lesson along with some high-level descriptive statistics of the lesson, detailed below. These statistics gave teachers objective feedback about various elements of their lessons. Once per term, the set of available

recordings was then analysed by the University of Melbourne using their teaching rubric and a personalised feedback report was written for each teacher.

The verbatim lesson transcripts were provided for download from the VC app in a Word or pdf file. The data dashboard was also accessible through the app and provided the teacher with key metrics about their teaching, which was calculated from an analysis of the captioned lesson. The reported metrics were percentage of talk time by teacher, number of teacher and student questions asked, and talk speed (words per minute). This information and feedback were presented visually in the form of annotated graphs. An example of the type of feedback provided is included in Appendix C.

The feedback report provided the teacher with feedback on their teaching over the five captioned lessons. Their lessons were coded against an evidence-based rubric of effective teaching practice, which identifies 16 practices or categories that were identified through a review of teacher effectiveness tools. During the intervention development phase, the rubric was refined over a series of iterations in which transcripts were practice-coded and the frequency counts were analysed statistically. An abridged version of the rubric is included in Appendix D. Teacher feedback was provided within a maximum of two weeks after the teaching took place. Through this process, areas of strength and areas for improvement were signalled to the teacher both in written and visual form. Trends over time were also provided and their practice was compared against teachers from similar schools (in terms of school type, region, and remoteness classification) to provide a context/benchmark for their performance.

Finally, teachers were provided with tailored suggestions for improvement, links to additional informational resources that might help put feedback into practice, and some prompts for discussion with their mentor.

4. **What (procedures)**: During the in-person training, teachers were introduced to the pedagogical model, the technological platform, and the ways in which they could use the feedback delivered to inform their practice. They also took part in a goal-setting exercise. This training comprised a mixture of presentations by the project team, interactive activities, and free time to explore the VC app and practice recording themselves speaking.

   During a one-hour webinar, mentor teachers were engaged in an interactive conversation about effective mentorship processes and were provided with hypothetical scenarios, resources, and ideas to support them to work with their mentee teachers in this project.

   To commence a captioned lesson over the course of the academic year, teachers established an audio link to the Ai-Media captioning centre via the VC app over a broadband internet connection via smartphone or tablet device and recorded their lesson. Trained staff then generated a verbatim transcript of the lesson from which the data dashboard was populated automatically using a series of algorithms. Within 24 hours of the lesson taking place, these statistics were released to teachers through the VC app.

   A team of trained research assistants from the University of Melbourne then coded the transcripts against an evidence-based rubric of effective teaching practice. An abridged version of this rubric can be found in Appendix D.

5. **Who (implementers)**: The intervention was delivered by a collaboration of three partners. Ai-Media provided the technological platform for the captioning of lessons, provided the verbatim lesson transcripts, and populated the data dashboard with teaching analytics. The University of Melbourne delivered the training package to participating teachers and mentors, conducted the in-depth coding of lesson transcripts, and generated the tailored feedback reports. SSAT were responsible for recruiting schools, checking that schools were using the technology correctly, and supporting them to do so effectively.

6. **How (mode of delivery)**: As detailed above, much of the intervention was delivered via technological platforms. Participating teachers attended a face-to-face training day, whilst their nominated mentors were encouraged to participate in an online webinar. Teachers recorded their lessons using the VC app, which could be accessed using a smartphone or tablet device. Their lesson transcript and data dashboard were also delivered through the app. After each academic term, the teachers received their detailed feedback report via email, which they were encouraged to review along with their transcripts and dashboard with their nominated mentor in a face-to-face session.

7. **Where (setting)**: The intervention was conducted in schools, except for the in-person training which was conducted off-site in a community venue.

8. **When and how much (dosage)**: Due to difficulties recruiting enough schools to maintain the necessary sample size according to the preliminary power calculations, the intervention was extended by an additional year, and thus divided into two phases. The first phase was divided into three blocks (a block referring to a deadline for submitting lessons and then receiving a feedback report), and within each block, teachers were expected to record and upload five one-hour lessons. The second phase of the evaluation was divided into two blocks, and within each block, teachers were expected to record and upload three one-hour lessons. The schools included in the evaluation remained largely constant across phases, barring slight discrepancies due to consent and data matching. For each hour they recorded and shared, teachers received a verbatim lesson transcript and access to a data dashboard, which gave an overview of the lessons they had uploaded. At the end of each block, teachers received an in-depth feedback report and a score on their teaching with tailored suggestions for improvement. They were encouraged to spend a minimum of one hour with their nominated mentor at the end of the block. This meant that teachers who participated across both phases of the evaluation were exposed to the intervention over two years (meaning that students in Year 5 during the first year of the evaluation *could have* continued to learn from a teacher using the VC app if they remained with a VC teacher during their Year 6), whereas teachers of students in Year 6 and/or joining the evaluation during phase two were exposed only for one year. Because the data does not allow us to identify which students were taught by teachers who were exposed to the intervention over two years, we do not know which students received a longer dosage and therefore cannot control for this. This is a limitation to our analysis as some students received longer dosage than others.

   During the planning stages, it was agreed that compliance with the intervention would be defined as having submitted ten records over the academic year and received at least two feedback reports from the University of Melbourne. Given the extension of the evaluation and the inclusion of phase two, we re-defined compliance as submitting either ten records across both phases, or, for teachers who joined in phase two, submitting five of the six possible records in phase two. We discuss compliance in further detail later in the report.

9. **Tailoring**: The intervention involved giving tailored feedback to teachers according to the statistics relating to their individual lessons, and therefore provided feedback that was adapted to the needs of each individual teacher. An example of the type of feedback provided can be found in Appendix C.

10. **Modifications**: School recruitment proved to be the greatest challenge during the project, compromising the sample size and our ability to detect impact. To mitigate the impacts of this problem on the study's power, we extended the trial by another year resulting in two years of data collection instead of one year, as originally planned.

11. **How well (planned)**: Online activity on the dashboard and website was measured; this acted as a robust measure of implementation. In particular, teachers' adherence to the intervention was tracked via the Visible Classroom app. Web analytics provided evidence of teachers' engagement with the intervention, including number of captioned lessons, number of reads of transcripts, and access rates of data dashboards. Teachers were also encouraged to record (via the app) when they met with their mentor teacher. Teachers with low levels of engagement with the intervention were contacted by SSAT for

additional support and troubleshooting. However, as previously discussed, students' exposure was not consistently tracked.

Evaluation objectives

The aim of the evaluation was to assess the impact of the Visible Classroom intervention on KS2 results in maths and reading for pupils who received the intervention in Year 5 or Year 6.

The primary research question was:

- Does the VC intervention increase the educational attainment of Year 5 and Year 6 students in their combined KS2 maths and reading scores, assessed at the end of Year 6?

The impact evaluation also aimed to address the following secondary research questions:

- Does the intervention have an impact on the maths and reading results for each year group independently?
- Does the intervention have an impact on the maths and reading results of free school meal (FSM) students?
- Does the intervention have an impact on the maths and reading results of students whose teachers score in the bottom third of the VC rubric, as measured at baseline?

The implementation and process evaluation (IPE) aimed at exploring how the intervention operated in practice, the barriers to, and conditions for, effective delivery of the intervention, and the related costs of implementation incurred by schools and teachers. Further details about the evaluation's objectives and approach can be found in the evaluation protocol here. The statistical analysis plan can be found here.

Ethics and trial registration

Ethical review was undertaken by the University of Melbourne Ethics Committee (approval reference number 1647332) and by BIT's External Ethical Advisory Panel. The project received ethical approval in June 2016 and the ethical approval letter can be found in Appendix L.

Before randomisation, schools received an information sheet about the trial and gave consent by signing an agreement to be part of the trial (Appendix E). The requirements for schools participating in the study were set out in this agreement, which all participating schools were required to sign. Because the intervention was delivered directly to teachers, consent was given by teachers when signing up to participate in the intervention. They were also given an information sheet (Appendix F) about the evaluation.

Students' parents (or legal guardians) made an informed decision regarding whether they wanted their children's Key Stage 2 results to be part of the analysis. An information leaflet was sent to parents along with the form that gave them the opportunity to withdraw their child's data from being used in analysis (Appendix G). This provided information on the aims of the research and the planned use of data to allow parents to make an informed decision about whether to agree to data sharing. The forms were returned by parents to their child's school, which then returned data to BIT only from students who did not opt out of their data being used for analysis.

The trial has been registered at: https://doi.org/10.1186/ISRCTN14774597.

Data protection

As noted above, schools, teachers and parents/carers were provided with an information sheet setting out the details of the evaluation prior to deciding whether to take part, with the requirements of participation detailed in the agreements. Parents/carers were also provided with a withdrawal form in case they did not wish their child's attainment outcome data to be included in the evaluation. These documents can be found in Appendices E, F, and G.

This project began in 2016 and the intervention was delivered prior to the introduction of GDPR in September 2016. At this time (pre-May 2018), parents were offered the opportunity to withdraw their child's data from analysis, which met the criteria for measures to protect participants and their data as set out in the Data Protection Act 1998. With the introduction of GDPR, BIT updated the data-handling and privacy policy to align with the new regulation. BIT justified the processing of the data under 'legitimate interests'. According to GDPR, there are three elements to the legitimate interests basis, and each element is justified below:

1. Identify a legitimate interest:

   This research has a positive social impact as it seeks to improve the educational outcomes of students. This work was intended to yield important outcomes for all key stakeholders. Teachers were expected to benefit from enhanced support and improved professional practice, students from improved attainment, and policymakers from an increased understanding of what works best to support children's attainment in reading and maths. In addition, parents and students may reasonably expect their attainment data to be used and processed for research purposes if the research seeks to benefit their child.

2. Show that the processing is necessary to achieve it:

   In order for BIT to assess whether the intervention has had its intended effect (that is, improvements in maths and reading outcomes) it is necessary to process individual level NPD data (the maths and reading attainment scores for students in the trial).

3. Balance it against the individual's interests, rights and freedoms:

   The rights of the individual were balanced by offering parents the opportunity to withdraw their child's data from analysis. Parents were also given further information and a point of contact for questions or concerns.

Only the members of the research team at the Behavioural Insights Team working on this project had access to the data. A three-year (maximum) license for access to the National Pupil Database data was requested. The research team sought the maximum length to ensure that there is enough time in case of any delays, which are common in research involving multiple agencies and stakeholders that all have different output requirements and publication deadlines.

The National Pupil Database is the data controller of the matched KS2 outcome data and the Behavioural Insights Team is the data processor. After the trial, the EEF will become the data controller for all trial data that we are authorised to release.

Project team

Development and delivery team

Janet Clinton (University of Melbourne): lead on intervention design, professor and study principal investigator.

John Hattie (University of Melbourne): lead on intervention design, professor and study co-investigator.

Anne-Marie Duguid (Schools, Students and Teachers Network): school recruitment lead.

Tony Abrahams (Ai-Media): developer of Visible Classroom platform.

Evaluation team

Dan Carr (Behavioural Insights Team): lead on impact evaluation design, statistical methods and project management.

James Lawrence (Behavioural Insights Team): project management and principal Investigator.

Matt Barnard (Behavioural Insights Team): support in analysis and report QA.

Hazel Wright (Behavioural Insights Team): project management support.

Jessica Heal (Behavioural Insights Team): implementation and process evaluation lead.

Juliane Wiese (Behavioural Insights Team): analysis and impact evaluation.

Sarah Breathnach (Behavioural Insights Team): analysis and impact evaluation.

Lauren Crouch (Behavioural Insights Team): implementation and process evaluation.

Michael Sanders (formerly the Behavioural Insights Team) was the Principle Investigator on the evaluation until December 2018. Tom Middleton (also formerly the Behavioural Insights Team) also provided support on analysis and impact evaluation.

Methods

Trial design

The evaluation utilised a two-arm cluster randomised controlled trial design comparing the Key Stage 2 maths and reading scores of students in the Visible Classroom programme to students in the control condition, with randomisation at school level. See Table **2** for a summary.

Table 2: Summary of the Visible Classroom programme evaluation

Trial type and number of arms		two-arm, cluster randomised
Unit of randomisation		School
Stratification variable(s) (if applicable)		Proportion of FSM pupils: above or below median KS1 Average Point Score: above or below median in 2010/2011 Offered entry on capped or uncapped basis. The last stratification covariate was added in mid-August 2016 as a result of a decision during school recruitment to begin capping the number of teachers per school who could be involved in the trial. This decision was made to avoid a cost overrun in the delivery team's budget. Capped schools may have selected which teachers could be involved in the trial on the basis of merit or some other element correlated with student attainment. As a result, it was included as a covariate in our stratification. The SSAT noted which were capped and which were uncapped.
Primary outcome	variable	Attainment in maths and reading (combined score) at end of Year 6, analysed for pupils receiving the intervention in Years 5 and 6 separately.
	measure (instrument, scale)	Combined KS2 maths (KS2_MATMRK) and reading (KS2_READMRK) scores for pupils receiving the intervention in Y5. Combined KS2 maths and reading scores for pupils receiving the intervention in Y6.

The Visible Classroom intervention occurred at the teacher level, however randomisation took place at the school level because it was felt that randomising at the teacher level would have entailed substantial risk of cross-contamination. Schools in the control group were expected to continue with 'business as usual'; teachers in these schools did not participate in the Visible Classroom intervention.

The trial was initially meant to run over the course of one academic year. However, due to difficulties recruiting enough schools and concerns around the statistical power of the trial, the evaluation was extended by an additional year during which teachers could use the VC app, and the process evaluation and intervention rollout occurred for a second year. This decision was taken without any knowledge of the results of the first year. One benefit of randomising at the school level was that classes would not be mixed up between treatment and control groups in the following year when the trial was extended, which could have happened had randomisation taken place at the teacher level.

Participant selection

The eligibility criteria for schools to participate were:

- upfront transfer of data to the evaluators including eligible student UPNs;
- a completed Memorandum of Understanding;
- agreement that teachers would complete a survey at the end of the trial period;
- inclusion of both Y5 and Y6 in the intervention, with a minimum of two teachers providing baseline data and attending the training session;
- teachers had access to a tablet or smartphone and sufficient internet connection to upload recordings; and
- the schools were not using Visible Learning plus, a similar intervention designed by the University of Melbourne Project Team.

Once a school was deemed eligible for the trial, there were no eligibility criteria for students other than that their parents did not withdraw their child's results for inclusion in the evaluation analysis.

The trial was conducted across 86 schools in England, although the original target was 140 schools. Of the initial 154 schools that applied to participate in the study, 138 were deemed eligible and were shortlisted. Of these, 26 schools failed to complete all the necessary stages of the recruitment process. This meant that 112 signed the Memorandum of Understanding. Prior to randomisation, 26 schools dropped out leaving 86 schools to randomise (see Figure 1 for further details).

The Schools, Students and Teachers Network (SSAT) was responsible for recruiting schools into the trial. Schools within chosen local authorities were invited to take part in the pilot study and evaluation. SSAT recruited from its wide network of schools and other contacts in England, including at an event in Dudley and through Nesta contacts. SSAT issued a call for expressions of interest from schools then held short phone conversations with interested schools to assess eligibility (criteria listed above). To start, 154 applications from schools were received; however, many of these were deemed excluded or withdrew (detailed in Figure 1). After recruitment, each school nominated teachers to participate. In most cases these comprised all the teachers for a whole year-group (such as all Year 5 teachers or all Year 6 teachers). Selected schools were intended to be representative of a range of potential users of the technology so that findings would be as generalisable as possible. This included schools in urban and rural areas, schools with higher and lower levels of technology experience, and those with more newly qualified teachers and those with more experienced teachers who might be less familiar with the use of technology in education. All schools were sent information packs and consent forms to be signed by all teachers involved in the project as well as information sheets for parents. These can be found in Appendices E, F, and G.

Outcome measures

Primary outcome

The primary outcome measure was a composite attainment score, combining maths and reading Key Stage 2 results, and analysed separately for Year 5 and 6 cohorts. Scores were combined by adding the raw scores together, which is the most straightforward means of combination and produces identical standardised effect sizes to more complex methods of combination. In addition, using an attainment outcome provided by the NPD reduced the likelihood of missing outcome data. The KS2 scores were obtained from the NPD, held by the Office for National Statistics (ONS).

Secondary outcomes

There were no secondary outcomes in this trial. Secondary analyses were conducted, as described below.

Sample size

The sample size deemed necessary at the trial protocol stage followed standard criteria on power (0.80) and level of significance (0.05). After discussions with the broader project team, the evaluation team estimated that 140 schools would realistically participate in the trial, with an average of 1.5 classes per year-group and an average class size of 28 students (42 students in each school cluster). The intracluster correlation coefficient was assumed to be 0.15 based on estimates found in the literature (Hedges and Hedberg, 2007). Because tracking was possible using Unique Pupil Numbers (UPNs), the team expected attrition to be minimal; however, the statistical analysis plan did set out a process in the instance that significant data was missing. Additionally, participants' KS1 results were incorporated as baseline control measures to increase power along with the variables used for blocking during randomisation. The team assumed a correlation coefficient of 0.7 between KS1 and KS2 results. On this basis, with the assumed 140 schools in the trial, the evaluation team estimated an MDES of 0.14.

Due to difficulties recruiting schools, the project team decided to extend the evaluation by one year. This meant that students in Year 5 during the first year of the evaluation who were then taught in Year 6 by teachers involved in the evaluation would be exposed to the treatment for two years rather than only one. Ultimately, 86 schools participated (86 in Year 5 and 85 in Year 6: this discrepancy was due to one school failing to return consent forms to use the Year 6 student data). In terms of the number of pupils ultimately enrolled in the trial, the total number was 7,230 (3,531 in Year 5 and 3,699 in Year 6), which is higher than the assumptions made in the original power calculation, but with a lower number of schools. Because randomisation occurred at the school (rather than pupil) level, this ultimately compromised the power of the evaluation. This meant the MDES at the point of randomisation was 0.18 for both Year 5 and Year 6 (see Table 10).

Due to imperfect matching to the NPD, the final analysis sample size was 6,934 students (3,337 from 83 schools in Year 5 and 3,597 from 85 schools in Year 6). The ICC ultimately calculated was 0.10 for Year 5 and 0.12 for Year 6, both lower than the predicted ICC of 0.15. This recovered some of the lost power from the reduced number of schools resulting in an MDES at the point of analysis of 0.16 for Year 5 and 0.17 for Year 6 (see Table 10).

For the subgroup of children eligible for FSM, at the time of writing the trial protocol we assumed seven pupils per cluster would be eligible for FSM (based on the 16.7% of students receiving FSM as determined by the average across all primary schools on the EduBase dataset). The assumption of seven FSM pupils per cluster gave an MDES of 0.18.

Ultimately the number of FSM-eligible children (as determined by the variable EverFSM_6_P in the NPD) was 2,576 (1,232 in Year 5 and 1344 in Year 6), with 15 per cluster in Year 5 and 16 per cluster in Year 6, which meant at analysis the MDES for this subgroup was 0.19 for Year 5 and 0.18 for Year 6. Note that the FSM indicator was only requested from the NPD at the point of analysis, so at randomisation the MDES reported in Table 10 for FSM reflects our analysis sample.

Randomisation

Randomisation followed recruitment of schools, including the signing of Memoranda of Understanding (MoU), distribution and receipt of consent forms by schools, and provision to BIT of child- and school-level data. The project delivery team had no role in the randomisation process.

Randomisation was conducted by a member of the BIT research team whose randomisation code, which can be found in Appendix J, ensured independence of allocation. Schools were stratified on the basis of school-level characteristics: proportion of FSM students (above vs below median), previous test scores (2010/2011 KS1 Average Point Score, above vs below median), and whether the school was recruited into the trial on a capped or uncapped basis. Information on FSM and KS1 scores was drawn from the DfE's Performance Tables rather than from data collected ourselves. The randomisation followed a two-stage process.

1. Schools were stratified on the basis of the variables listed above. This stratification resulted in eight blocks and schools were randomised within each block. The purpose of this blocking was to improve cross-arm

comparability of schools and also to increase precision of estimates. Block sizes are described in Table **3** below.

2. A random number was generated within each block and the subsamples were split into two groups of equal size to ensure that school FSM proportion, KS1 test score proportion, and capped/uncapped schools were balanced across trial arms.

A total of 86 schools were recruited. For Year 5, 42 were randomised to control and 44 to treatment; for Year 6, 42 were randomised to control and 43 to treatment (one Year 6 cohort was excluded because the school did not return consent withdrawal forms from these students). In Table 3 and Table 4 we set out the breakdown across stratification variables for Years 5 and 6, respectively. In a small number of cases, data required for stratification was missing (four schools are missing FSM data and three schools are missing KS1 data). These schools were placed into their own strata. In Year 5 this included five schools and in Year 6 this included four schools.

Table 3 and Table 4 note the final number of schools randomised by stratification variable.

Note that the analysis was not undertaken blind to randomisation. We do not feel that this biases the results because the evaluation team was not responsible for delivery of the intervention; thus, knowing the randomisation assignments would not impact or bias any interactions with the randomised schools.

Table 3: Randomisation strata by batch—Year 5

		Equal to or below-median FSM	Above-median FSM
KS1 point score: above median	Capped entry	4	2
	Uncapped entry	24	6
KS1 point score: equal to or below median	Capped entry	1	4
	Uncapped entry	10	27

Table 4: Randomisation strata by batch—Year 6

		Equal to or below-median FSM	Above-median FSM
KS1 point score: above median	Capped entry	4	2
	Uncapped entry	25	7
KS1 point score: equal to or below median	Capped entry	1	4
	Uncapped entry	11	27

Statistical analysis

Primary analysis

As set out in the Statistical Analysis Plan (SAP), impacts were estimated for the primary analysis using a least squares linear model with treatment arm indicators, stratum indicators (that is, whether the school was above or below the median FSM proportion, whether the school was above or below the median KS1 score, and whether the school was offered entry on a capped or uncapped basis), and KS1 scores as a measure of prior attainment. This analysis was conducted separately by year level.

As this is a school-level randomised controlled trial, inference was based on standard errors adjusted for school-level clustering using Stata's 'cluster' option to allow for correlation of pupil outcomes within schools.

The estimated impacts captured the effect of intention to treat (ITT) and are reported with 95% confidence intervals. Intra-cluster correlations (ICCs) are also reported. The model estimated for the primary analysis was:

Equation 1: $Y_{ijt} = a + \beta_1 Treat_{ij} + \beta_2 Y_{ij(t-1)} + \beta_3 \gamma_{ij} + \varepsilon_{ijt}$

where i are individuals and j are schools, Y_{ijt} is our endline KS2 score (with the value of this at t - 1 being the KS1 score for the same student), Treat was our school-level treatment indicator, γ_{ij} being a set of stratum dummies, where each stratum corresponds to a unique combination of the stratification variables, and ε_{ijt} being an error term. Errors are clustered at the school level (j). Our primary intention to treat outcome was recovered from the estimate of β_1. This model was not altered according to the significance of any variables included (that is, all variables were retained in the model regardless of whether they are statistically significant)—including the vector of blocking variables (γ_{ij}).

In line with the approach set out in the SAP, estimates are presented as effect sizes, calculated using the Hedges' g formula, expressing the estimated effect as represented by the regression coefficient relative to the total unadjusted outcome variance in the sample.

Secondary analysis

We repeated the primary analysis using individual scores for the KS2 attainment measures (that is, KS2 maths and KS2 reading scores separately), with analysis for Years 5 and 6 conducted separately. This offered insight into which domain drove the primary result as well as an overall robustness check on the main analysis.

Subgroup analysis

Impacts were estimated using the same approach for the subgroup of pupils eligible for FSM (using the variable EVERFSM_6_P available from the NPD in line with EEF guidance). The analysis was run separately for the FSM subgroup as specified in the SAP.

Missing data

The evaluation protocol specified that in the event that there was greater than 5% rate of missing outcome data, we would fit a regression where the outcome is the missingness of Y_{ijt}, and where the explanatory variables are student-level prior attainment, gender, and free school meals status, and characteristics of the school including its Ofsted rating. The level of missingness for KS2 outcome data was 2.8% for Year 5 and 2.0% for Year 6 and thus it was not deemed necessary to perform sensitivity analysis and impute the missing KS2 scores.

The SAP did specify that additional sensitivity analysis would be conducted for pupils with KS2 attainment data but no KS1 attainment data. Regarding KS1 data, 3.4% was missing in Year 5 and 5.1% in Year 6. In these instances, KS1 performance was imputed using the school's mean KS1 score for the primary analysis.

Compliance with intervention

The evaluation team and the University of Melbourne agreed and documented in the SAP that in order to be considered compliant for analysis of the primary outcome, a teacher must have submitted at least ten recordings over the academic year (out of a maximum 15). However, given cost and timing considerations that changed with the extension of the programme, the second phase allowed for a maximum of six submissions. It was decided, therefore, that compliance would be defined as submitting ten recordings overall—as this was the original expectation for an evaluation lasting one year—or five recordings in the second phase (as there were far fewer opportunities for submitting recordings than in the first phase). Thus, it was necessary to allow teachers who joined only in the second phase to be able to meet the bar of compliance. Compliance data was collected by the University of Melbourne, which kept track of each teacher's submissions in each phase of the programme.

We used Complier Average Causal Effect (CACE) analysis in addition to the primary ITT to estimate intervention effects on treated children (Gerber and Green, 2012). We estimate the CACE using two-stage least-squares (2SLS) regression by estimating a (first stage) model of compliance using the binary measure of compliance described above. The predicted values from the first stage are then used in the estimation of a (reduced form) model of our outcome measure Y_{ijt}. In other respects, the specification remains the same as the primary outcome ITT model. We conducted this analysis using the ivregress functionality of Stata to make necessary adjustments to standard errors (which are also clustered at school level) due to the instrumental variables approach. We note the need for caution in generalising the results of this analysis given the likely endogeneity of recording submissions and motivation. This means that it is possible that more highly motivated teachers are more likely to comply with the intervention, which means that our sample of outcome data in the CACE analysis is inherently biased and not fully representative of the broader teacher population.

Additional analyses

We conducted a descriptive dose response analysis in order to understand the effect of recording additional lessons on student outcomes. Our treatment variable indicates the number of lessons recorded and uploaded by the teacher, and is conducted only on student outcomes of those students whose teachers uploaded recordings of their lessons.

As a supplement to the evaluation, teachers were sent an online survey focused on their own current teaching practices and experiences with VC, both before and after the intervention was rolled out. Pre-surveys took place in June 2016 for phase one and September 2017 for phase two; post-surveys were administered in July 2017 for phase one and April 2018 for phase two. We conducted analysis using this Survey of Teacher Practice to understand whether and how self-reported changes in teaching practice moderate any effect on attainment. To do this, we chose three broad question types to explore: frequency of using various teaching strategies, extent to which feedback is valued by teacher, and extent to which feedback is valued by school. The survey was also intended to promote self-evaluative practices as it asks teachers to reflect on how often they use these teaching strategies and also to identify their three priority areas of improvement. Each question type contained a number of sub-questions. We averaged teacher responses to sub-questions within each type and regressed the primary outcome measure on responses as well as the other covariates from the primary analysis model.

Implementation and process evaluation

A process evaluation was conducted alongside the quantitative component of the study described above. The primary aim of the process evaluation was to understand the barriers and facilitators to implementation and, in particular, the factors affecting fidelity and engagement. The secondary aim of the process component was to identify the factors influencing the effectiveness of the intervention. The primary approach to data collection was semi-structured interviews with teachers and mentors. In addition, training sessions were observed and the

materials provided to teachers were reviewed. The original design included an online survey but, as discussed below, challenges in recruiting the sample means that findings from the survey are not included in this report. Further details about process evaluation methodology are set out below.

Semi-structured interviews

A total of 26 interviews with teachers and mentors were conducted across years one and two of the intervention. Participants were informed that the interview was anonymous, that they could withdraw at any time, and that they did not have to answer any questions they did not want to. The delivery team were not told which schools were visited.

All interview data was analysed using thematic analysis (Braun and Clarke, 2006). Participants' responses were identified, coded line by line, and analysed using the computer-assisted qualitative data analysis software, Dedoose. Codes were regularly organised into groups, revised, and refined until they formed the themes and sub-themes presented in this report.

Sampling

Teachers and mentors were selected based on a purposive sampling approach. Unlike sampling for a quantitative survey, purposive sampling does not aim to generate a statistically representative sample. Instead, participants are selected with the aim of capturing the range and diversity of views and experiences across the whole population. For this study, we aimed to sample 20 of each—teacher and mentor, mentor only, and teacher only—across both years. We also aimed to sample respondents that reflected high, medium, and low engagement with the intervention. Engagement level was based on number of lesson submissions and is explained in further detail below. Due to sampling difficulties, we could not recruit a balanced number of participants across those categories. This was primarily due to potential participants either dropping out of the programme entirely or not responding to our recruitment emails and calls. However, the sample did include a range of respondent types (mentors, teachers, and teacher/mentors) and a range of levels of engagement. A summary of the characteristics of the achieved sample is included in the tables below and a more detailed breakdown is included in Appendix H.

Table 5: Sample information by role type

	Year 1, 2017	Year 2, 2018
Total interview participants	20	6
Teacher and mentor	1	1
Mentor	6	0
Teacher	13	5

Table 6: Sample information by engagement

Interviewee engagement	Year 1, 2017	Year 2, 2018
High	18	3
Medium	2	1
Low	0	2

Intervention compliance

Fidelity and compliance were measured through submissions of lesson recordings via the application which were received from Ai-Media. This also allowed for dosage of the intervention and the subsequent identification of low- and high-adherence teachers. Critical features of the intervention, with a sliding scale of engagement, are outlined below. It considers the rounds of lessons and recordings.

To ensure the integrity of the data was maintained, it was important to define what, for teachers, constituted 'receipt of the intervention'. This was defined as receiving a minimum of two detailed feedback reports from UoM.

Table 7: Engagement with, and compliance to, the intervention—phase one

Submit lessons	Round 1	Round 2	Round 3	Total submissions
High	5	4–5	4–5	13–15
Medium	3–4	3–4	3–4	7–12
Low	0–2	0–2	0–2	0–6

In year two, the requirements on submission targets were reduced and only two rounds of three recordings were conducted. Therefore, for year two, the engagement was revised, as shown in Table **8**.

Table 8: Engagement with, and compliance to, the intervention—phase two

Submit lessons	Round 1	Round 2	Total
High	3	3	5–6
Medium	2	2	3–4
Low	0–1	0–1	0–2

This data was used for sampling purposes in the IPE as well as for CACE analysis in the impact evaluation.

Surveys

Data from the interviews was used to inform the design of the brief post-intervention survey, which was distributed online to teachers who participated in Visible Classroom during 2017 and 2018. The survey, conducted in the summer of 2018, was designed to capture satisfaction, perceptions of usefulness, and general feedback on the intervention. When this survey was initially sent out to teachers by the evaluation team, no teachers responded. To improve the response rate, respondents of the survey were offered entry into a prize draw to win a £50 Amazon voucher. However, despite this incentive and three email reminders being sent, only four teachers responded. We have not included any analysis of the survey as, due to the low response rate, we do not feel findings would provide a robust indication of the prevalence of the views of teachers involved in Visible Classroom. It is possible that the poor response rate may be a reflection of teacher engagement with the intervention more generally as we also struggled to sample for the qualitative component of the evaluation.

Costs

Cost information was collected by the University of Melbourne. This information captured the cost of teacher training, recruitment of Research Assistants (RAs), transcript analysis by RAs, and any additional materials or support provided to schools.

For mentors and teachers, we used usage data from the platform to determine the amount of time a teacher typically spent engaging with the intervention. We also requested information via surveys on the amount of time spent in relation to the intervention, for example, training, discussions with mentors, time added to lessons due to recording equipment, if any, and time spent on the website dashboard.

Timeline

The Table **9** below outlines the timeline for the evaluation.

It is important to note that the draft report was submitted to the EEF over a year after the end of the evaluation. This was due to a series of delays that we experienced when applying for NPD data. The delays were a result of significant changes that were made to the NPD request process to ensure better handling and safeguarding of pupil data. As part of this, the Secure Research Service (SRS) was introduced. This service makes NPD data available to researchers through secure portals at the ONS. All researchers wishing to access this data must undergo training to become registered SRS researchers. The introduction of the SRS meant that BIT staff had to undergo training and complete all analysis onsite at the ONS. This caused significant delays to the evaluation.

Table 9: Timeline

Date	Activity
Apr–Sep 2016	Recruitment of schools
Mar–Apr 2016	Ethical approval
Apr–Sep 2016	Eligibility interviews with schools
Sep 2016	Randomisation
Sep–Oct 2016	Training with schools
Sep 2016–Jun 2018	Full rollout of intervention
May–Jul 2017	Year 1 Process Evaluation
May–Jul 2018	Year 2 Process Evaluation
Nov 2018	Initial NPD request submitted
Dec 2018	Initial NPD request rejected
Jan 2019	NPD request re-submitted
Jan–Mar 2019	Application goes through NPD Governance

Mar 2019	NPD request accepted
Mar–Jul 2019	NPD merge datasets
Jul–Oct 2019	Data accessed through the SRS
Oct 2019	First draft of Evaluation full report

Impact evaluation results

At the time of analysis, the MDES in Year 5 was 0.16, reflecting 83 schools and 3,337 children, and 0.17 in Year 6, reflecting 85 schools and 3,597 children. The NPD was unable to match the KS2 records of 296 students, resulting in the attrition of approximately 4% of the sample. As all data was obtained via the NPD (rather than being collected directly through students) there was minimal attrition between the randomisation and analysis stages. This low level of attrition and missingness of outcome data did not change the MDES drastically from the one calculated at the time of randomisation. In fact, the lower-than-predicted ICC in the analysis sample improved power for the evaluation. The flow diagram below (Figure **1**, page 24) provides details of participant flow through each stage of the evaluation, while Table 10 (page 25) provides details of the MDES at the different stages of the trial.

Figure 1: Participant flow diagram

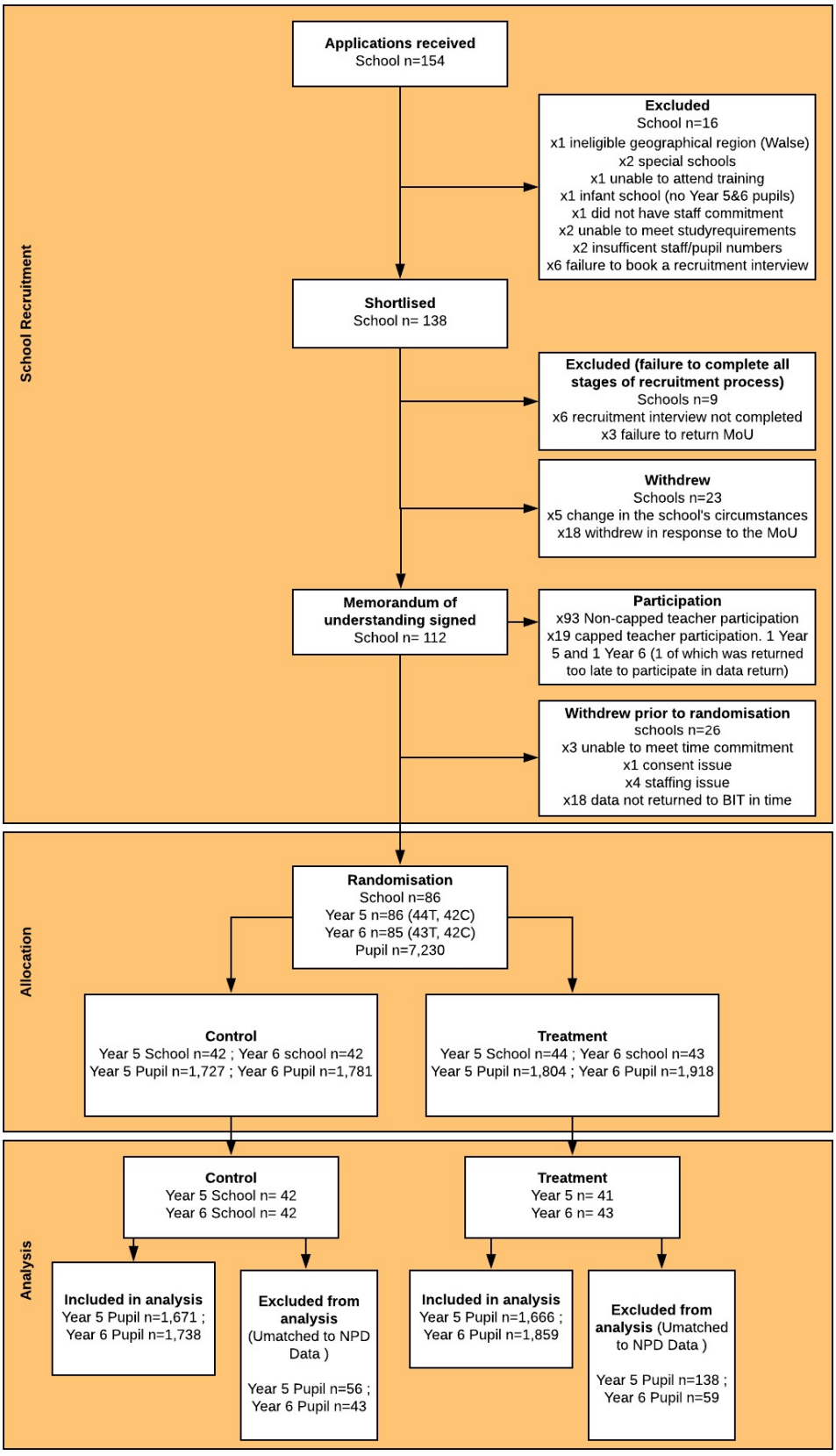


Table 10: Minimum detectable effect size at different stages, for Years 5 and 6 separately

		Protocol (Y5 + Y6)		Randomisation (Y5)		Randomisation (Y6)		Analysis (Y5)		Analysis (Y6)	
		Overall	FSM	Overall	FSM[1]	Overall	FSM[2]	Overall	FSM	Overall	FSM
MDES		0.14	0.18	0.18	0.19	0.18	0.18	0.16	0.19	0.17	0.18
Pre-test/ post-test correlations	Level 1 (pupil)	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
	Level 2 (class)										
	Level 3 (school)										
Intracluster correlations (ICCs)	Level 2 (class)										
	Level 3 (school)	0.15	0.15	0.15	0.15	0.15	0.15	0.10	0.12	0.12	0.12
Alpha		0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
One-sided or two-sided?		2	2	2	2	2	2	2	2	2	2
Average cluster size		42	7	42	15	41	16	40	15	42	16
Number of schools	Inter-vention	70	70	41	40	43	43	41	40	43	43
	Control	70	70	42	41	42	41	42	41	42	41
	Total	140	140	83	81	85	84	83	81	85	84
Number of pupils	Inter-vention	2940	491	1804	636	1918	719	1666	636	1859	719
	Control	2940	491	1727	596	1781	625	1671	596	1738	625
	Total	5880	982	3531	1232	3699	1344	3337	1232	3597	1344

Attrition

As predicted in the SAP, student-level attrition in this trial was minimal due to outcomes and baseline measures being available from the NPD. In total, only 4% of the sample could not be matched NPD data. Because this level of attrition for the primary outcome measure is less than the pre-agreed threshold of 5%, we did not carry out the main missing data sensitivity analysis specified in the SAP.

[1] Please note that these figures represent the FSM calculations at point of analysis, given that we only have access to FSM data with the analysis sample.

However, the level of missingness of KS1 outcome data was 3.4% for Year 5 and 5.1% for Year 6 respectively. Therefore, we conducted additional sensitivity analysis for pupils with KS2 attainment data but no KS1 attainment data by imputing the schools' mean KS1 scores.

Pupil and school characteristics

Table 11 and Table 12 present school- and pupil-level characteristics for the intervention and control groups in Years 5 and 6, respectively. The school-level characteristics remain the same at randomisation and analysis because there was no attrition at the school level; however, pupil-level characteristics reflect the analysis sample only because we only have access to FSM eligibility, gender, KS1 baseline scores, and age for the analysis sample (which is slightly smaller than the randomisation sample due to some attrition from the ONS data match). The sample was balanced at baseline on all of the characteristics we were able to observe. There were no statistically significant differences by treatment arm in the distribution of schools by Ofsted rating, school type, or urban/rural location. There were also no statistically significant differences by trial arm in the composition of pupils within schools (based on the number of pupils in schools and the percentage of FSM pupils). There is a small absolute difference in KS1 results between treatment and control groups with those in the control group scoring slightly higher than those in the treatment group. There is also a slightly higher percentage of FSM pupils in the treatment group. However, neither of these differences were statistically significant. When we consider characteristics at the pupil level, we find that the sample is balanced in respect to FSM eligibility, gender, KS1 baseline score, and age.

We also report the average individual pre-test scores in

Table 11 and Table 12 for the analysis sample in Years 5 and 6. This cannot be done using the baseline sample because the scores were provided to us via NPD matching that excluded those students that could not be successfully matched. We present histograms showing the distribution of these scores in Figure A1 in Appendix I.

Table 11: School- and pupil-level characteristics in Year 5

School-level[2] [3] (categorical)	Intervention group		Control group		National avg, primary schools[4]
	n/N (missing)	Count (%)	n/N (missing)	Count (%)	Count (%)
Ofsted rating					
Outstanding	7/44 (6)	7 (15.9%)	6/42 (4)	6 (14.3%)	2,924/16,777 (17.4%)
Good	24/44 (6)	24 (54.5%)	27/42 (4)	27 (64.3%)	11,508/16,777 (68.6%)

[2] As reported in Edubase; data accessed June 2019.
[3] These figures are the same at randomisation and analysis because no schools dropped out.
[4] As reported by Gov.uk Official Statistics. This column includes schools in England as at 31 March 2019. The granularity of the data does not allow us to break this down beyond the primary school level; therefore, we do not repeat this column for Year 6 in Table 13 as Years 5 and 6 both belong to primary schools. In addition, we only report those figures that were provided to us by the DfE.

Requires improvement	7/44 (6)	7 (15.9%)	4/42 (4)	4 (9.5%)	1,769/16,777 (10.5%)	
Special measures	0/41 (6)	0 (0.0%)	1/42 (4)	1 (2.4%)	474/16,777 (2.8%)	
School type						
Academy converter	9/44 (0)	9 (20.5%)	3/42 (0)	3 (7.1%)	3,806/16,777 (22.7%)	
Academy sponsor-led	7/44 (0)	7 (15.9%)	6/42 (0)	6 (14.3%)	1,453/16,777 (8.7%)	
Community school	18/44 (0)	18 (40.9%)	20/42 (0)	20 (47.6%)	6,231/16,777 (37.1%)	
Foundation school	2/44 (0)	2 (4.5%)	4/42 (0)	4 (9.5%)	570/16,777 (3.4%)	
Free schools	1/44 (0)	1 (2.3%)	0/42 (0)	0 (0.0%)	170/16,777 (1.0%)	
Voluntary aided school	7/44 (0)	7 (15.9%)	6/42 (0)	6 (14.3%)	2,763/16,777 (16.5%)	
Voluntary controlled school	0/44 (0)	0 (0.0%)	3/42 (0)	3 (7.1%)	1,784/16,777 (10.6%)	
Urban / Rural						
Urban	38/44 (0)	38 (86.4%)	36/42 (0)	36 (85.7%)	-	
Rural	6/44 (0)	6 (13.6%)	6/42 (0)	6 (14.3%)	-	
School-level[5] (continuous)	n (missing)	Mean (SD)	n (missing)	Mean (SD)	Mean (SD)	
Number of pupils	37 (7)	371.8 (162.6)	37 (5)	354.2 (170.8)	281.2 (161.9)	
Percentage eligible for FSM	37 (7)	20.8 (15.0)	37 (5)	15.5 (11.7)	-	

[5] These figures are the same at randomisation and analysis because no schools dropped out.

Pupil-level[6] (categorical)	n/N (missing)	Count (%)	n/N (missing)	Count (%)		
Eligible for FSM	334/1666 (2)	334 (20.0%)	279/1671 (8)	279 (16.7%)	-	
Female	830/1666 (0)	830 (49.8%)	821/1671 (0)	821 (49.1%)	-	
Pupil-level[7] (continuous)	n (missing)	Mean (SD)	n (missing)	Mean (SD)		Effect Size
KS1 Baseline score	1621 (45)	33.1 (6.5)	1604 (67)	33.5 (6.2)	-	0.06
Age in September 2018 (months)	1665 (1)	138.5 (3.6)	1666 (5)	138.6 (3.6)	-	

Table 12: Baseline comparison Year 6

School-level[8][9] (categorical)	Intervention group		Control group	
	n/N (missing)	Count (%)	n/N (missing)	Count (%)
Ofsted rating				
Outstanding	7/43 (6)	7 (16.3%)	6/42 (4)	6 (14.3%)
Good	23/43 (6)	23 (53.5%)	27/42 (4)	27 (64.3%)
Requires improvement	7/43 (6)	7 (16.3%)	4/42 (4)	4 (9.5%)
Special measures	0/43 (6)	0 (0.0%)	1/42 (4)	1 (2.4%)
School type				

[6] These figures represent the analysis sample only as this data was only available to us at the ONS for those students whose UPNs matched to these characteristics.
[7] These figures represent the analysis sample only as this data was only available to us at the ONS, for those students whose UPNs matched to these characteristics
[8] As reported in Edubase; data accessed June 2019.
[9] These figures are the same at randomisation and analysis because no schools dropped out.

Academy converter	9/43 (0)	9 (20.9%)	3/42 (0)	3 (7.1%)
Academy sponsor-led	7/43 (0)	7 (16.3%)	6/42 (0)	6 (14.3%)
Community school	18/43 (0)	18 (41.9%)	20/42 (0)	20 (47.6%)
Foundation school	2/43 (0)	2 (4.7%)	4/42 (0)	4 (9.5%)
Voluntary aided school	7/43 (0)	7 (16.3%)	6/42 (0)	6 (14.3%)
Voluntary controlled school	0/43 (0)	0 (0.0%)	3/42 (0)	3 (7.1%)
Urban / Rural				
Urban	37/43 (0)	37 (86.0%)	36/42 (0)	36 (85.7%)
Rural	6/43 (0)	6 (14.0%)	6/42 (0)	6 (14.3%)
School-level[10] (continuous)	n (missing)	Mean (SD)	n (missing)	Mean (SD)
Number of pupils	36 (7)	377.2 (161.6)	37 (5)	354.2 (170.8)
Percentage eligible for FSM	36 (7)	20.7 (15.2)	37 (5)	15.5 (11.7)
Pupil-level[11] (categorical)	n/N (missing)	Count (%)	n/N (missing)	Count (%)
Eligible for FSM	346/1859 (0)	346 (18.6%)	292/1738 (0)	292 (16.8%)
Female	935/1859 (0)	935 (50.3%)	867/1738 (0)	867 (49.9%)

[10] These figures are the same at randomisation and analysis because no schools dropped out.
[11] These figures represent the analysis sample only as this data was only available to us at the ONS for those students whose UPNs matched to these characteristics.

Pupil-level[12] (continuous)	n (missing)	Mean (SD)	n (missing)	Mean (SD)	Effect Size
KS1 Baseline score	1781 (78)	32.8 (6.6)	1644 (94)	33.1 (6.4)	0.05
Age in September 2018 (months)	1859 (0)	150.2 (3.8)	1738 (0)	150.1 (4.0)	

Table **13** presents absolute standardised differences for the continuous variables specified in the statistical analysis plan, KS1 baseline scores and age, for the analysis sample, split by year. As expected, the KS1 baseline scores and age are the same.

Table 13: Absolute standardised differences for selected characteristics, analysis sample for Years 5 and 6

		Intervention group Mean (SD)	Control group Mean (SD)	Absolute standardised difference[13]
Year 5	KS1 baseline score: maths	16.4 (3.3)	16.6 (3.2)	0.06
	KS1 baseline score: reading	16.7 (3.6)	16.9 (3.5)	0.06
	Age (months)	138.5 (3.6)	138.6 (3.6)	0.03
Year 6	KS1 baseline scores: maths	16.3 (3.2)	16.4 (3.3)	0.03
	KS1 baseline scores: reading	16.5 (3.7)	16.7 (3.6)	0.06
	Age (months)	150.2 (3.8)	150.1 (4.0)	0.03

Finally, for characteristics that are categorical variables (gender and FSM eligibility), we ran a regression of the characteristics on treatment allocation and found that there was no significant difference between intervention group and control group along these characteristics (Table **14**).

[12] These figures represent the analysis sample only as this data was only available to us at the ONS for those students whose UPNs matched to these characteristics.
[13] This was calculated as the absolute difference between intervention and control means, divided by the control group standard deviation.

Table 14: Regressions for comparison of categorical variables, analysis sample

		Female	Ever FSM
Year 5	Allocation	0.007 (0.018)	0.024 (0.051)
	N	3,337	3,332
Year 6	Allocation	0.004 (0.016)	0.026 (0.048)
	N	3,597	3,556

Note: standard errors based on school-level clustered standard errors reported in parentheses. Statistical significance indicated as follows: * p < 0.05; ** p < 0.01; *** p < 0.001.

Outcomes and analysis

Analysis code can be found in Appendix J.

Primary analysis

The analysis indicates a non-statistically significant impact of the programme on the primary outcome measure, with a slightly higher control mean than treatment mean. Barring statistical significance of the effect, the difference in treatment and control outcomes is equivalent to one month's less progress for the treatment arm.

Table 15 summarises the results of the primary analysis. In both Years 5 and 6, comparison of the mean scores for the treatment and control groups indicates a lower mean on the composite KS2 outcome measure (the primary outcome) in the intervention group (109.32 in Year 5; 106.68 in Year 6) compared with the mean in the control group (113.79 in Year 5; 109.71 in Year 6). The distribution of scores across treatment and control groups, which are slightly skewed to the right, are presented in

Figure **2** and Figure 3 (page 33). The parameters for the effect size calculation are reported in Table 16.

Table 15: Primary analysis

	Raw means				Effect size		
	Intervention group		Control group				
Outcome	N (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Hedges g (95% CI)	p-value
KS2 maths and reading, Year 5	1666 (0)	109.32 (107.86, 110.78)	1671 (0)	113.79 (112.34, 115.23)	3,337 (1666; 1671)	-0.09 (-0.22, 0.04)	0.191
KS2 maths and reading, Year 6	1859 (0)	106.68 (105.22, 108.13)	1738 (0)	109.71 (108.21, 111.22)	3,597 (1859; 1738)	-0.05 (-0.19, 0.09)	0.55

Table 16: Effect size estimation

Outcome	Unadjusted differences in raw means	Adjusted differences in means[14]	Intervention group		Control group		Pooled variance	Population variance (if available)
			n (missing)	Variance of outcome	n (missing)	Variance of outcome		
KS2 maths and reading, Year 5	-4.47	-2.64	1666 (0)	922.34	1671 (0)	907.24	914.78	-
KS2 maths and reading, Year 6	-3.04	-1.38	1859 (0)	1028.40	1738 (0)	1020.50	1024.59	-

[14] This refers to the coefficient of the Treatment dummy in the regression.

Figure 2: Histogram of KS2 outcomes—Year 5

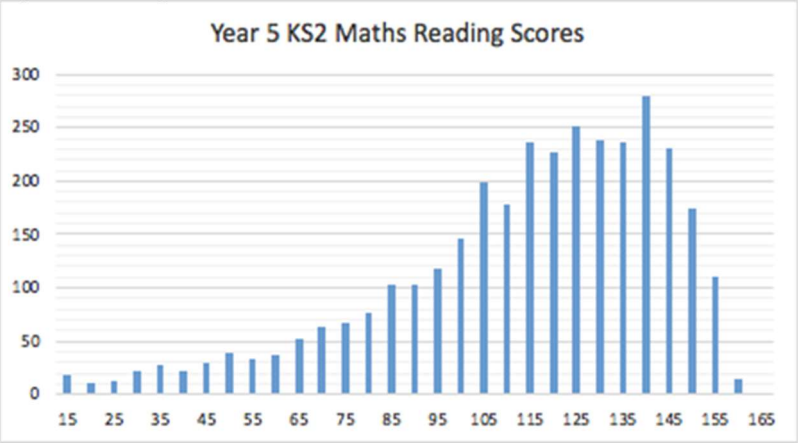
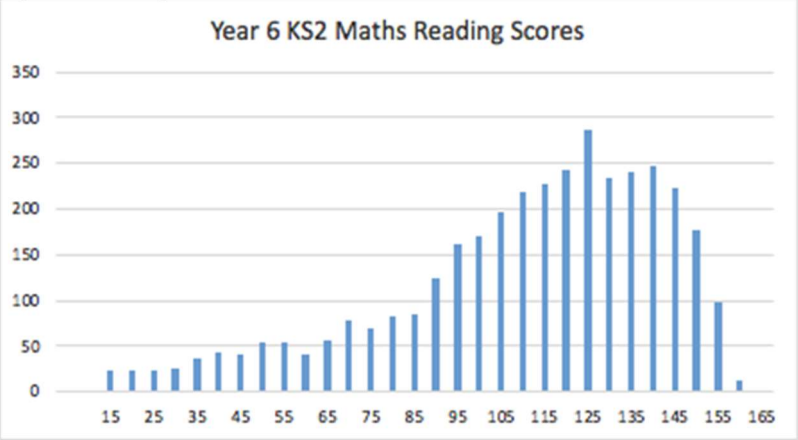


Figure 3: Histogram of KS2 outcomes—Year 6



Secondary analysis

As specified in the trial protocol, we repeat the primary analysis using individual scores for the KS2 attainment measures (that is, KS2 maths and KS2 reading scores separately). Again, the analysis for Years 5 and 6 is conducted separately. In all specifications, the mean scores in the treatment group are smaller than the mean scores in the control group. The effects of the treatment on KS2 maths raw scores were not statistically significant in either Year 5 and 6. However, the negative impact of the treatment on KS2 reading raw scores was statistically significant for Year 5 and equivalent to two months less progress. The impact of the treatment on KS2 reading scores in Year 6 was not statistically significant. The results of these analyses are presented in Table 17.

Table 17: Secondary analysis—Years 5 and 6

	Raw means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Hedges g (95% CI)	p-value
KS2 Maths, Year 5	1666 (0)	76.64 (75.53, 77.74)	1671 (0)	79.52 (78.42, 80.63)	3337 (1666;1671)	-0.07 (-0.20, 0.07)	0.34
KS2 Reading, Year 5	1666 (0)	32.68 (32.24, 33.13)	1671 (0)	34.26 (33.84, 34.69)	3337 (1666;1671)	-0.13 (-0.25, -0.01)	0.04
KS2 Maths, Year 6	1859 (0)	75.77 (74.66 76.87)	1738 (0)	77.64 (76.50, 78.79)	3597 (1859; 1738)	-0.03 (-0.18, 0.11)	0.73
KS2 Reading, Year 6	1859 (0)	30.91 (30.46, 31.36)	1738 (0)	32.07 (31.62, 32.52)	3597 (1859; 1738)	-0.08 (-0.20, 0.04)	0.209

Subgroup analysis—FSM pupils

In Year 5, the sample of pupils eligible for FSM again shows a negative effect size consistent with the findings in the primary analysis with the full sample; however, the magnitude of the effect size is closer to zero, is not statistically significant, and equates to no effect on progress. In Year 6, we find that the unadjusted control mean of the KS2 outcome is higher than the unadjusted treatment mean; however, when controlling for the strata and baseline scores in our regression, we find a positive, non-statistically-significant impact on KS2 scores. Table 18 reports the results of estimating the primary model for the FSM subgroup and Table 19 presents the underlying parameters for the effect size calculation.

Table 18: Impact on primary outcome—FSM pupils

	Raw means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; control)	Hedges g (95% CI)	p-value

KS2 maths and reading, Year 5	636 (0)	98.72 (96.17, 101.27)	596 (0)	105.49 (102.96, 108.03)	1232 (636; 596)	-0.03 (-0.23, 0.17)	0.77
KS2 maths and reading, Year 6	719 (0)	98.68 (96.18, 101.18)	625 (0)	100.69 (98.07, 103.31)	1344 (719; 625)	0.03 (-0.13, 0.20)	0.71

Table 19: Effect size estimation—FSM pupils

Outcome	Unadjusted differences in raw means	Adjusted differences in means[14]	Intervention group		Control group		Pooled variance	Population variance (if available)
			n (missing)	Variance of outcome	n (missing)	Variance of outcome		
KS2 maths and reading, Year 5	-6.77	-0.94	636 (0)	1072.23	596 (0)	994.11	1034.44	-
KS2 maths and reading, Year 6	-2.01	1.15	719 (0)	1170.08	625 (0)	1114.81	1144.38	-

Missing data

The SAP specified that for students with KS2 attainment data but no KS1 scores, KS1 performance would be imputed using the school's mean KS1 score. In this section we report the results of the missing data analyses specified in the SAP. Because the level of missingness for KS2 outcome data was below 5% in both Years 5 and 6, it was not deemed necessary to perform sensitivity analysis and impute the missing KS2 scores.

We did, however, perform additional sensitivity analysis for pupils with KS2 attainment data but no KS1 attainment data as an additional robustness check to our primary analysis, which includes imputed KS1 baseline scores. On this basis we still find no statistically significant effect of the intervention on the combined primary KS2 maths and reading outcome, as shown in Table 20. The findings are consistent with the primary analysis on the full sample (which includes imputed KS1 scores) at randomisation.

We run the same specification as above for our secondary analysis and find outcomes consistent with the full sample at randomisation: there is a statistically significant negative impact of the intervention on KS2 reading scores for Year 5 and no statistically significant impact in any of the other specifications (Table 21).

Table 20: Primary outcome excluding missing KS1 scores

	KS2 maths and reading, Year 5	KS2 maths and reading, Year 6
Treatment	-2.717 (2.004)	-1.489 (2.267)

N	3225	3425

Note: standard errors based on school-level clustered standard errors reported in parentheses. Models also control for blocking dummies of strata used in randomisation and KS1 scores. Statistical significance indicated as follows: * p < 0.05; ** p < 0.01; *** p < 0.001.

Table 21: Secondary outcomes excluding missing KS1 scores

	KS2 maths, Year 5	KS2 reading, Year 5	KS2 maths, Year 6	KS2 reading, Year 6
Treatment	-1.546 (1.563)	-1.171* (0.539)	-0.756 (1.780)	-0.733 (0.593)
N	3225	3225	3425	3425

Note: standard errors based on school-level clustered standard errors reported in parentheses. Models also control for blocking dummies of strata used in randomisation and KS1 scores. Statistical significance indicated as follows: * p < 0.05; ** p < 0.01; *** p < 0.001.

Compliance with the intervention

In this section we present the results of analyses assessing the impact of compliance on the primary outcome. We find that there is no statistically significant impact of being a complier of the treatment on the KS2 combined maths and reading score. We estimated the CACE using two-stage least-squares (2SLS) regression by estimating a (first stage) model of compliance using the binary measure of compliance described earlier in the report. The predicted values from the first stage were then used in the reduced form model of our outcome measures. This was done separately for Year 5 and Year 6.

One hundred and thirty-seven teachers (81%) were considered to be compliant—approximately the same percentage as the number of intervention pupils identified as having compliant teachers (80%). For both Year 5 and Year 6, the instrument, treatment allocation, was strong, with F statistics of 34.4 and 50.9, respectively. In addition, the R^2 in both years is relatively high (0.56 in Year 5 and 0.57 in Year 6), which suggests that the treatment allocation explains a high proportion of variation in the compliance variable. This provides strong evidence that pupils in control schools were not taught by teachers receiving the intervention.

In sum, the reduced-form model did not find a statistically significant impact of being a complier of the treatment on the KS2 combined maths and reading score. Indeed, the coefficients are negative, meaning that combined maths and reading scores were lower for those in the treatment group than those in the control group; however, this difference is not statistically significant from zero. This is in line with the results of the primary analysis.

Table 22: Second stage regression model of complier on KS2 maths and reading scores—Year 5

	KS2 maths and reading
Complier	-3.780 (2.855)
N	3337

Note: standard errors based on school-level clustered standard errors reported in parentheses. Models also control for blocking dummies of strata used in randomisation and imputed KS1 scores. Statistical significance indicated as follows: * p < 0.05; ** p < 0.01; *** p < 0.001.

Table 23: Second stage regression model of complier on KS2 maths and reading scores—Year 6

	KS2 maths and reading
Complier	-1.965 (3.232)

N	3597

Note: standard errors based on school-level clustered standard errors reported in parentheses. Models also control for blocking dummies of strata used in randomisation and imputed KS1 scores. Statistical significance indicated as follows: * p < 0.05; ** p < 0.01; *** p < 0.001.

Additional analyses

Descriptive dose response

A descriptive dose response analysis was conducted in which the primary outcome measure was regressed on total number of recordings submitted including only those teachers who submitted a recording of their lessons. The results show that an additional submission of a recorded lesson results in decreased KS2 combined maths and reading scores, in both Years 5 and 6. However, these results were not statistically significant.

Table 24: Descriptive dose response

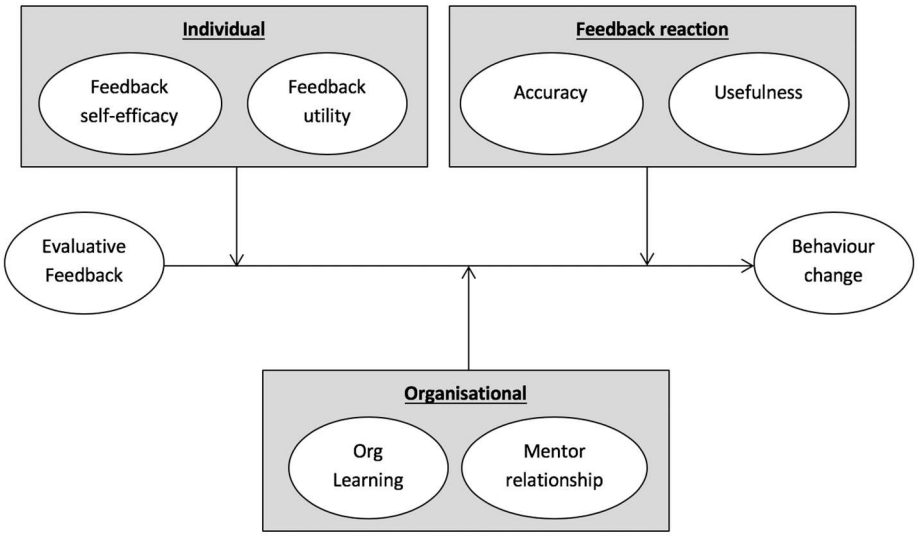
	KS2 maths and reading, Year 5	KS2 maths and reading, Year 6
Total number of lessons recorded and submitted	-0.279 (0.170)	-0.171 (0.186)
N	1387	1579

Note: standard errors based on school-level clustered standard errors reported in parentheses. Models also control for blocking dummies of strata used in randomisation and KS1 scores. Statistical significance indicated as follows: * p < 0.05; ** p < 0.01; *** p < 0.001.

Survey of Teacher Practice

Finally, the results of the analysis on the Survey of Teacher Practice developed by the University of Melbourne are presented below. We received pre-intervention and post-intervention survey responses from intervention schools. There were 57 surveys from Year 5 teachers and 66 surveys from Year 6 teachers. As can be seen in Figure **4** below, the survey contains three main themes which relate to (i) the individual, (ii) reaction to feedback, and (iii) the organisation.

Figure 4: Model of the Survey of Teacher Practice



The survey was used to understand whether and how self-reported changes in teaching practice moderate any effect on attainment. The survey was collected pre-intervention and post-intervention in each phase.15 Survey items were developed based on the literature and focused on the frequency of using teaching strategies, how much teachers value feedback, and the extent to which feedback is valued in their particular school. Exploring how often teachers use certain teaching strategies—and their perception of how they or their school value feedback—examines the self-reflective, evaluative mindset of teachers and whether this is related to, or is a mechanism for, teacher practice change to occur. This research explored the hypothesis that if teachers (and indeed their mentor teachers) have positive beliefs around the utility of feedback for practice improvement (having an evaluative mindset) then they will more readily engage with their feedback reports, have more reflective and productive conversations with their mentor, and improve in key areas of practice. A copy of this survey and further details on its development can be found in Appendix K of this document.

The results show that none of the question responses suggest that any changes in teaching practice moderate any effect on attainment except for the specification of teacher valuing of feedback in Year 5. This specification suggests that teachers reporting that they highly value feedback has a negative impact on KS2 outcomes.

Table 25: Regression of frequency of using teacher strategies on primary student outcomes

	KS2 maths and reading, Year 5	KS2 maths and reading, Year 6
Frequency of using teacher strategies	-25.763 (16.250)	4.577 (5.011)
N	321	482

Note: standard errors based on school-level clustered standard errors reported in parentheses. Models also control for blocking dummies of strata used in randomisation, KS1 scores, and the relevant question's pre-intervention response. Statistical significance indicated as follows: * p < 0.05; ** p < 0.01; *** p < 0.001.

Table 26: Regression of teacher's valuing of feedback on primary student outcomes

	KS2 maths and reading, Year 5	KS2 maths and reading, Year 6
Teacher's valuing of feedback	-12.319* (4.869)	8.277 (4.954)
N	321	482

Note: standard errors based on school-level clustered standard errors reported in parentheses. Models also control for blocking dummies of strata used in randomisation, KS1 scores, and the relevant question's pre-intervention response. Statistical significance indicated as follows: * p < 0.05; ** p < 0.01; *** p < 0.001.

Table 27: Regression of school's valuing of feedback (reported by teacher) on primary student outcomes

	KS2 maths and reading, Year 5	KS2 maths and reading, Year 6
School's valuing of feedback (reported by teacher)	-7.159 (9.852)	4.654 (3.332)
N	321	482

Note: standard errors based on school-level clustered standard errors reported in parentheses. Models also control for blocking dummies of strata used in randomisation, KS1 scores, and the relevant question's pre-intervention response. Statistical significance indicated as follows: * p < 0.05; ** p < 0.01; *** p < 0.001.

[15] The phase one pre-survey was collected in June 2016, and the post-survey in July 2017. The phase two pre-survey was collected in September 2017, the post post-survey was collected in April 2018.

Cost

Schools participating in the intervention received one day of training before the start of the intervention in 2016 for both Year 5 and Year 6 teachers. Aside from this, each teacher was left to record and upload lessons in order to receive a tailored feedback report.

Most of the delivery costs that schools incurred related to training, staff cover, and, in some cases, buying equipment to record the lessons. All these costs were accrued in the first year; there were no additional costs in subsequent years for teachers who had already had the training (see Table 28).

The total cost per pupil over three years is based on the number of pupils who received the intervention—that is, on the basis of three teachers per school (on average) with an average class size of 27 (see Table 28). Training costs per participating teacher works out at 850 AUD (£490) and assumes they receive lesson transcripts, have access to dashboard analytics, and receive one feedback report.16 They also receive a pre-prepared user guide and can request technical assistance and support from the Ai-Media team when required. In our assumptions of cumulative costs (

Table **29**), we assume teachers will remain engaged in the intervention each year. The largest costs are fixed costs that occur in the first year.

Table 28: Individual school cost of delivering Visible Classroom

Item	Type of cost	Average Cost	Total cost over 3 years	Total cost per pupil per year over 3 years
Materials	Recording devices, microphones, device cases	£169	£169	£1
Staff travel and subsistence	Expenses	£366	£366	£2
Training fee	Training	£1,470	£4,410	£18
Total		£2,005	£4,945	£21

Source: University of Melbourne: teacher survey and teacher interviews.

Table 29: Cumulative costs of Visible Classroom

Cumulative costs	£	Number of pupils	Per pupil £	Cost Rating

[16] When used in a commercial setting outside the scope of this evaluation, Ai-Media charges approximately 360 AUD for transcripts, analytics and feedback report for a five-lesson series.

Year 1	2,604	81	32	£
Year 2	4,074	162	25	£
Year 3	5,544	243	23	£

Source: University of Melbourne: teacher survey and teacher interviews.

As outlined above, the intervention becomes better value for money the more pupils that are exposed (assuming that staff do not leave and need retraining). This means the cost of the programme would be 'very low' by EEF definitions.

Start-up costs and ongoing costs in Year 1

As highlighted above, assuming that no new staff join and need training and materials, there are no additional expenses to the school beyond the start-up costs.

Table 30: Costs of the intervention in Year 1

	Per School (£)	Per Pupil (£)
Training cost	1,470	18
Start-up costs	1,134	14
Running costs	0	0
Total cost Y1	2,604	32

Source: University of Melbourne: teacher survey and teacher interviews.

School size

We have calculated the average number of pupils by dividing the number of pupils in the trial by the number of schools in the trial and by the average number of teachers trained per school (three). Given that this programme was trialled across two year groups, the minimum number of staff per school that could have been trained is two whilst some larger schools had nine teachers trained. The size of the school has significant implications for the costs of the intervention, as it is per teacher.

Technology expenses

Technology expenses—such as the purchase of iPads—were greater in schools whose policies prohibited the use of personal iPads and phones in class. Roughly 25% of the sample that returned the cost survey purchased additional equipment.

Staff time

The intervention requires one day of training for each participating teacher. This time cost is magnified as the majority of cost survey respondents highlighted that they paid for cover for the training day, which was attended by at least two teachers per school. The reported hours to embed the intervention and cover staff absence ranged from two days to four and a half days, with an overall average of three days.

Table 31: Average staff time required to deliver the intervention

Description	Average Days (8 hour days)	Average Hours
Embedding the intervention (training, recording)	1.3	11
Planning time (mentoring; reflection; planning changes)	2.15	17
Other (sending emails, administration)	0.5	0.38

Source: teacher survey; teacher interviews.

Implementation and process evaluation results

This section sets out the findings of the implementation and process evaluation. The first subsection below sets out findings related to factors affecting implementation, while the second subsection focuses on the factors influencing the effectiveness of the intervention.

Factors affecting implementation

The evaluation identified four key themes:

- Initial engagement—teachers expressed some initial apprehension about receiving feedback and some felt self-conscious about being observed, but motivation to take part was bolstered by interest in research into the approach and self development.
- Training and buy-in—there was a lack of clarity among some about logistical aspects of the programme, including difficulties in getting to grips with the technology.
- Ongoing participation and support—some teachers found it challenging to continue to engage in the programme or implement recommended changes due to professional and personal time pressures.
- Application and feedback style—feedback was valued because of its clarity and objectivity but, in some cases, teachers would have liked more detail and to receive it more immediately.

The themes are discuss in more detail below.

Initial engagement

Teachers and mentors reported some initial concerns about engaging in the project. The notion of receiving feedback made some feel apprehensive with one describing it as a 'daunting experience'. The project was described as initially feeling similar to a formal teaching observation. However, these concerns did not necessarily discourage participants from engaging in the first place.

> *'I was a little bit wary because I know from some of the feedback from the other teachers it was to do with pace of voice and questioning and things. So, it was something that I was very wary about going into it' (T15MP).*

One impact of participants' initial concerns was that some reported delaying recording or uploading lessons due to feeling self-conscious about being observed and assessed on particular elements of their practice.

> *'It was normally me not wanting to upload it because it was like, "Oh my god, is this a good enough lesson?"' (T15MP).*

> *'The facilitation of the lessons the first time was quite tricky. I think because I was very conscious, and I'm used to recording myself using video and then reflecting on it. But because I had a key purpose in mind, for example, which was to control the speed of my talk etc., I was very conscious throughout the first one' (T19LPM).*

In response to teachers' apprehension about receiving feedback, senior leadership team, teachers who had taken part in previous years, and Visible Classroom staff provided encouragement that helped assuage initial worries. Findings from the interviews suggested that the factors that motivated teachers to engage included having a genuine interest in Visible Learning research (which for some complemented academic study they were undertaking) and being eager to engage in an opportunity for self-development; this was particularly influential for newly qualified teachers (NQTs) aiming to achieve Qualified Teacher Status.[17] Building on teachers' individual motivations to take part also reduced nerves surrounding the feedback for some while increasing the focus on the positive outcomes of engaging.

[17] Qualified Teacher Status is the assessment during the NQT teaching year to receive a license to teach.

> *'We had read so much of Hattie's work. We were already quite steeped in his thinking in terms of Visible Learning, so that was pretty much integral to everything that we were doing. When the opportunity came up ... it was too good an opportunity to miss' (T19LPM).*

> *'I'm doing a masters at the moment ... and one of the areas that I have chosen to focus on is looking at the success criteria, and they have gone hand in hand' (T16HP).*

For some teachers, feelings of self-consciousness eased throughout the delivery of the project as they became increasingly familiar and comfortable with the process.

> *'You got used to that and you became less conscious in the fact that you were recording yourself' (T21HP).*

Training and encouraging buy-in from teachers

A number of teachers and SLTs described being confused about key messages relating to the intervention. For instance, some were not clear about how long the recordings needed to be, what to do when the class is interrupted, and the exact role of the mentor. In one school, the role of the mentor was misinterpreted as being intended as a tool for SLT to survey teaching staff. There were also members of SLTs who were surprised to find they had to fund supply teachers to cover staff who attended the training day. Within this context, there were questions about the need to be present at the training as well, and whether there were more cost-effective methods of disseminating the same information.

> *'I wouldn't say that we found [the training] hugely beneficial. It took us a few hours to get [there] on the train and taxis. The initial induction maybe we could have done without because [we haven't] gained like a wealth of knowledge from that … I don't think that we got a lot from it. We didn't come out of it thinking, "right, we know exactly what to do". It gave us lots of statistics on what works and what doesn't work, which a PowerPoint could have been shared and maybe a video' (SLT6M).*

At the same time, other senior leaders felt that it was essential to attend the training. Interestingly, teachers reported learning how to use the app in training and none reported difficulty using the app when delivering the project.

> *'We discussed how the app works ... initially we may have uploaded a recording and we were just given email instructions about that' (T16HP).*

Initial issues recording lessons was a common experience among participants and several teachers lost lessons they had recorded. Difficulties with technology included running out of storage space on phones, devices only recording parts of lessons, recordings having poor sound quality, and having a limited number of recording devices to share among staff. It was also reported that some teachers were told that using their phones in the classroom would contravene the school's safeguarding policy. It was clear that a lack of technological knowledge and understanding lay behind some teachers' difficulties with the recording equipment.

> *'It was partly my fault by having it below a table when I was talking to a group and it didn't pick me up. I think the only thing recording like that is it was very dependent upon the equipment you had ... there was one where I recorded a session and it didn't upload the full hour; it only uploaded the first couple of minutes. That happened a couple of times' (T6HP).*

> *'I think the first recording I did it didn't work, because when you press record and I turned the screen off and then it just turned the recording off. When you record the lesson, the phone is unlocked. That gets touched in moving around, or it knocks into things or catching me. That can quite often cause it to pause or to stop. If you lock the phone ... then it stops the recording. You had thought that you had done an hour recording, and then realised that it has paused after two minutes and you have lost the lesson' (T15MP).*

In one case, challenges in dealing with the technology led a teacher to drop out of the intervention (as reported by another teacher at their school). However, the majority of participants interviewed as part of this study felt it was a minor hindrance and expected some challenges when relying on technology and the internet to deliver this type of intervention.

> *'It has been great, and there have been challenges. Any time you involve the internet and [wireless technology] and trying to upload and things like that has been challenging' (T21HP).*

Contacting the Schools, Students and Teachers (SSAT) network for support and advice resulted in one teacher problem-solving unexpected technological difficulties that was affecting their ability to record lessons.

> *'I emailed the chat about a couple of questions and said that a couple of times the recordings stopped or paused or got lost' (T16HP).*

Over time, participants felt that the Visible Classroom app became easy to use and did not appear to encounter long-term difficulties uploading recordings or receiving and viewing the feedback.

> *'It was just record your lesson, upload it, and here is your feedback. It was very simple, and that's the thing I like about the app is that it's fairly easy and there is not a lot of set-up to it. It's login, record, and upload. I think it was quite clear' (T16HP).*

Ongoing participation and support

Maintaining engagement in a year-round programme can be difficult on top of the regular daily demands that teachers face. In light of this, many teachers benefitted from reminders to record their lessons; however, a number of teachers would have found it helpful to receive reminders to use the app.

Specific work and home life commitments were mentioned by teachers that affected their ability to regularly record lessons. For example, several teachers mentioned that the third session in the late Easter and early summer term was particularly difficult due to SATs. This was due to English and maths lessons predominantly consisting of revision lessons that teachers didn't feel were appropriate to record. In addition, some teachers worried that certain pedagogic methods, like group work, were not suitable lessons to record.

> *'Then as you get further on and it's my first year with Year 6 and it's more pressured, and you end up doing whatever you can. Also, the timing for the last one was really unfortunate. It was revision for SATs and it was like I'm not teaching … I wasn't mindfully implementing things, and I was more focused on revision or post-SAT stuff' (T3HP).*

Teachers also highlighted that if home life commitments were approaching (such as taking paternity leave) they felt less able to implement changes from the feedback due to a lack of time and energy to reflect on their practice and introduce changes.

> *'One of the main things was my classroom manner. I know we want open-ended questioning and looking at the vocabulary and I absolutely had those resources and if I had the time … trying to manage a small child, and look after my Year 6 this year has been very tricky, and lots of different juggling acts. I wouldn't say my words per minutes have really improved, and I think that is just down to me not necessarily changing my practice and that's a shame' (T16HP).*

There was also a sense of confusion for some about whether or not there were strict criteria regarding which lessons they should record, with several teachers mentioning that the type of lessons they delivered restricted their ability to record. Opening up to a wider variety of subject lessons that can be recorded may alleviate some of this pressure.

> *'For me, I was restricted to just doing English and maths lessons. In the afternoon in Year 6, for me particularly, it was revision. Especially from towards the end of February up until May, all of that time is revision' (T14HP).*

Teachers also felt that receiving reminders of upcoming recording submission deadlines supported their continued engagement in Visible Classroom by helping them to remember and prioritise recording sessions during busy periods. These reminders were particularly important as several teachers explained that their motivation to participate waned through the year.

> *'It was also difficult to remember to do it, especially it was the autumn term and you're so busy, your brain isn't in it. You are just so busy. It's not the top of your priority' (T20LP).*

> *'As the year went on, I got less and less motivated by the project, so I didn't use the feedback as much' (T4HP).*

Teachers did not receive reminders to use the app. Reminders via email or the app could be effective, as the Visible Classroom app was identified by participants as a useful and simple tool for delivering the project:

> *'It was difficult to always do it and maybe more reminders … either by email or via the app to remind you. That could send you a message couldn't it, it could send you a reminder' (T20LP).*

Although some teachers reported they had adequate time to submit recordings, others recommended increasing the time between deadlines. One teacher reported that the ability to reschedule and be flexible in his submission deadlines enabled his continued engagement in Visible Classroom alongside taking paternity leave.

> *'I had to get the recordings done though … at the time I wasn't really at the school, so I had to reschedule that and push it back a bit later' (T16HP).*

Some teachers found the strict deadlines difficult to meet. Issues meeting the recording deadlines were compounded by additional pressure in both a teacher's personal or professional life, such as sickness, Ofsted pressures, job shares, and school events. One teacher in particular mentioned that taking part in Visible Classroom felt like a significant extra pressure.

> *'The other thing that was really difficult was one of the sets of recordings was around SAT time and that was really hard for us, and we weren't teaching normal lessons. That was really hard' (T2HPM).*

> *'The only issue that we had is if you were off sick for a week during that period it became a bit more difficult to try and cram them into one week' (T5HP).*

The flexibility of Visible Classroom staff and SSAT support, in particular allowing deadlines to be flexible, enabled teachers to engage in the programme who might not otherwise have been able to.

Application and feedback style

The structure and style of the feedback delivered to teachers were crucial to how teachers understood their progress and areas for development. Most teachers felt the feedback was clear, comprehensive, and useful and in some cases described the process of identifying their strengths and areas for improvement as 'fun'. Feedback being presented via the use of visuals and graphs was also experienced as a powerful method of communicating areas of change.

> *'I think because you were looking at a visual rather than a set of numbers. Seeing it as a bar chart is quite powerful and whether that is a positive or negative it's still powerful' (T19LPM).*

Many teachers valued the objective style of the feedback and felt that it was evidence-based and data-driven, often contrasting it to the kind of feedback they typically received. The fact that the feedback was seen as objective helped them accept it as a more accurate portrayal of their performance and therefore take it on board more willingly.

> *'It was less offensive because it wasn't someone's opinion. If it was negative you knew it was easier to get on with and sort. Often your instinct is to disagree, or you genuinely think that they're wrong, so it was quite good at being data fuelled' (T20LP).*

Providing a more neutral source of information about their progress helped some teachers apply their feedback to classroom practice.

> *'I quite liked that you couldn't disagree with it. It was programmed in and you either had these problems or you hadn't. You had words speaking in an optimum range for children to understand all your words. So it's easier to take it rather than someone's opinion' (T20LP).*

In some cases, teachers said they would like more information regarding certain aspects of the feedback, such as the reading scales, in order to improve their understanding and subsequent ability to make changes to these areas.

> *'I didn't really understand the Flesch reading scale right at the end, the comprehension that you use was the only bit I was unsure about. I think it was to do with words that we were using and whether it was relevant to our cohort that we're teaching. It would be nice to have more information on that' (T15MP).*

Some teachers felt they could benefit from the feedback even more if they received more immediate feedback to enact changes and improvements in the classroom more quickly. This was because they felt their memory of the lesson would be fresher, and therefore the feedback would have felt more relevant.

> *'Perhaps a little bit quicker… instant feedback on your weaknesses would be great, maybe two weaknesses and two strengths, instant feedback, as in words per minute. You know—how they did those word cards, something like that' (T20LP).*

For teachers, comparability of feedback was also important. They explained that making changes to their practice based on the feedback was difficult when the feedback directly compared different classes of pupils or subject lessons between the first and second year of participation.

> *'I know from round one I had a really different class. So trying to compare those two, because in the graphs for this round you have both phase one and two compared and that was hard because it was completely different children and completely different needs. Sometimes when you get the report back about the behavioural questions, it's difficult to put that into context because all three of those lessons can look quite different … that was the number of questions that I perhaps asked in that English lesson and trying to compare that with a maths lesson is really tricky because they are very different lessons' (T16HP).*

Additional resources—such as using examples of good teaching practice in television programmes that are aimed at similarly-aged children to their pupils—made the feedback more applicable and assisted teachers in making changes to their classroom practice.

> *'The talk time didn't come down and I need to work on that, but I think—and this is fed back at the celebration event and we were talking about needing to see what that looked like. Their reference was the television, actually, and they said they actually shared some children's television clips, which was aimed at particularly young children in which the adults were speaking at 120 words a minute and what that looked like. Then using that as a model because you are aware of it, you are able to deliver it' (T21HP).*

However, despite resources being available, a number of participants said that it would have been useful to have had better signposting to informational resources more directly relevant to feedback. They also said that having additional information about the length of time it may take to engage with the resources may better support teachers to implement actual changes and practical techniques:

> *'You could have tailored examples, and I remember once or twice there were examples on how to help the weaknesses that have been flagged up. But you could link and show articles on teacher practice to your weakness. If you have certain weaknesses that round then you get an email on things to read up on and have a look at. But if speaking too fast is recurrently coming up as one of your weaknesses, then yes, having articles pinged into your account about why that's a problem that would be really useful' (T20LP).*

Interviews with teachers suggest that not all of them made equal use of the materials and resources to help them apply the feedback. More scaffolding and expectations around the use of resources may contribute to a more consistent experience across all participants.

Factors influencing effectiveness

The process evaluation identified three factors that appear to influence the effectiveness of the intervention.

1. *Teachers' mindset*

   This was influenced by whether teachers' previous experience of feedback was positive or negative. However, the ease with which teachers could engage with the programme supported its use and also

prompted informal discussions about feedback with colleagues. The individualised and private nature of the programme also helped embed it in teacher's practice.

2. *The nature of the mentor-teacher relationship*

   In some schools discussion between mentors and teachers were included as part of the timetable, but where discussions were solely informal, teachers said they would have valued more formal and structured sessions. Relationships between mentors and teachers that were non-hierarchical and collaborative were experienced positively and this type of relationship was reinforced when teachers were both mentors and participants.

3. *The school environment*

   The programme appeared to fit best in schools where there was an existing culture of learning and development. Where this culture did not exist, teachers described feeling that the programme was used as a monitoring and performance management tool rather than a learning and development opportunity.

These three factors are discussed in more detail below.

Teachers' mindset

Teachers' mindset is crucial if they are to engage with and apply feedback to their practice. There were several mechanisms that emerged from the qualitative data that shed light on how this mindset shift may have been initiated. One important influence was prior experience of feedback, which held particularly strong connotations for a number of interviewees.

> *'Observations can be a bit of a torture and intimidating. I think by doing something like this, that takes the intimidation right out. Also I think having it as a recording makes it less stressful as well' (T5HP).*

The ease with which teachers could engage in feedback helped it to become part of their regular routine. Teachers stated that minimal extra lesson preparation and planning was required to stay engaged with recording and uploading the sessions throughout the project. The simplicity of engaging in the programme is important to ensure long-term engagement, especially as teachers are working within a climate of increased accountability and pressure.

> *'It was handy that even just the setting up of it and walking around with it on the lanyard it takes a few seconds. It doesn't take time to do that, so it was really helpful. I just did what I would normally do' (T18HP).*

The ease with which teachers could engage in feedback also led to an increased number of informal conversations about feedback, which were not tied to pressurised formal observations.

> *'I discussed with the teacher who works next door quite often ... and didn't particularly have any worries about it. It wasn't a case of, "What did you get and what did you get." But it was more like, "I had my feedback today", and "How did you get on?" and that sort of discussion ... quite informal' (T16HP).*

The individualised nature of the programme may also have contributed to reformulating the concept of feedback for teachers from something that was provided by others to something they could control and take ownership over.

> *'The pressure is different and you are not being held accountable to it; you are purely being accountable to yourself' (T3HP).*

Providing teachers with the opportunity to take ownership over their feedback supported personal goal-setting. Some teachers felt the application was a useful tool to set their own targets and monitor their own progress towards achieving these goals:

> *'I've used the app to track analytics and use that in how much I've improved ... so I've set that as one of my targets and that was good to see that over time, I felt that had improved' (T16HP).*

These individualised targets were motivating for several teachers, acting as a goal to work towards. Visually displaying and recognising progress they had made in altering their practice acted as a motivator to continue to apply the feedback to their practice:

> *'It was evident in the next feedback I got and it did say that you have made a clear effort to improve this. Then the next steps from a development point were less and less each time' (T14HP).*

These evident targets and goals also heightened teacher awareness of their own practice and enabled them to break down their practice so they could more effectively target on a particular element to improve.

> *'They would give you one thing to work on, so actually it's really valuable because it picks apart all the different bits of your teaching' (T2HPM).*

The individualised and private nature of the feedback may have helped shift teacher mindset, moving away from highly-planned observed feedback to one which is more embedded in their practice:

> *'I think my awareness of the way that I'm questioning the students has definitely increased, and my awareness of the importance of those questions … I did find myself making a conscious effort to ask more open questions as opposed to closed questions to try to extend learning. So for instance, asking a child a certain question and making sure that they are digging as deep into that answer as possible. Rather than just accepting the first answer and trying to make them explain further into it and make them understand their logic behind their answer' (T1MP).*

The mentor-teacher relationship

A positive mentor experience appeared to be an influential factor in how Visible Classroom was perceived and implemented by teachers.

> *'From my point of view it was helpful to have an experienced colleague to talk with, who would say the disadvantages of doing that are … and the advantages of doing that are … and discuss that together. Some of them are obvious, but others, it's like, "Is that a positive or a negative?"' (T1MP).*

> *'It's always good to verbalise to somebody, so it forces you to read through your report and verbalise and it helped to have somebody to bounce ideas around because we then picked a target to try and focus on … coaching each other is part of our school anyway so that was fine' (T2HPM).*

Similarly to the inconsistent use of materials and feedback by teachers, however, the way the mentor-teacher relationship was organised varied across case study schools. Mentors could be peers, the Visible Classroom project lead, or members of the school's SLT. In some schools, mentors and teachers were matched using random allocation, but the matching could also be determined by year group or on the basis of the experience of the teacher. A further approach was for the mentoring to be a group exercise rather than one-to-one.

> *'We did it collectively. We decided at the start that we were all going into this and whatever we said in those meetings was between the three of us' (T19LPM).*

In some schools, regular discussions between the mentor and teacher were included as part of timetables. However, informal 'catch-ups' between teachers and their mentors were also valued as they provided an opportunity to reflect on day-to-day changes that were being made to teaching practice:

> *'So, although we had the mentoring sessions, we had almost daily check-ins. Never a day went by without us talking about where we were up to, what impact we made, or the changes in the classroom' (T19LPM).*

In schools where the contact between mentors and teachers consisted solely of informal, unplanned catch-ups, the opportunity to have more formal meetings would have been welcomed along with more structure to guide the sessions.

> *'I think maybe some proforma so that there was some control over that session ... could have helped. For example, perhaps saying, "Look at this first, now have a look at that. How does this work? How does it compare with someone else?" That might have helped the discussion' (T19LPM).*

The pressures on school staff and the importance of having directed and productive mentor meetings were also highlighted.

> *'Ultimately, it's just the time to be able to get together and sit down and have those meetings without it being extra to your senior management meeting, the curriculum staff meeting, or your team meeting. It's finding those windows of time when we are very busy to sit down and have that meaningful conversation' (T21HP).*

A relationship that was non-hierarchical, collaborative, honest, professional, and non-judgemental was felt to be most effective in creating a space where feedback and areas for improvement could be openly discussed.

> *'The mentor sessions were really enjoyable and just great professional dialogues about the feedback, and how they were finding the ongoing project, and about the meetings' (T21HP).*

Participating in the project as both a teacher and mentor increased mentors' understanding of the process and made for more powerful target- and goal-setting as it fostered a sense of shared experience between the mentor and teacher.

> *'I had two people on the trial—I was the mentor for both ... that was great, but I wasn't living and breathing it in the sense that I hadn't tried it myself or understood the impact it had when you read some reflective feedback on your own practice. When I had the opportunity in the second year of the trial to be part of the process, I continued in the mentoring role, but also as part of the project. I think that helped because when we were looking at the next steps based on the analytics that we had, I was doing similar things, so I think it became more powerful' (T19LPM).*

> *'I don't know how beneficial that was because my mentor wasn't involved with the study. She knew about it and I didn't know if that was a good thing that she didn't know and she was looking at it with a freshness. It might have been better if I had a mentor from Year 5 who looked at it' (T3HP).*

School environment

A whole-school culture of learning and development played an important role in facilitating the Visible Classroom programme as it meant there was a natural synergy between the programme and the existing context. A culture of learning and development was evidenced by the existence of a cycle of continued professional development for teaching staff, support staff, and SLT. This helped ensure that the feedback from Visible Classrooms was implemented as intended.

> *'That is a feature of the school where all the senior leaders are and do have [significant] teaching commitments … and this is great continued professional development. Everything is triangulated, and this makes us a very reflective school. We are a school where the teachers are constantly engaged in learning. We are a school that very much signs up to the collective efficacy idea, which I know has been described in Hattie's work as a very powerful influence' (T21HP).*

In some cases, the school culture was seen as being strongly influenced by the approach of the senior leadership team and in particular the headteacher.

> *'I think we are a school that has quite a visionary head, and I think we all sign up to what she brings and believes' (T21HP).*

Teachers at schools with positive learning and development cultures connected the approach to continuing professional development for staff to the approach they had to teaching and the desire to promote a positive attitude to life-long learning among pupils.

> *'There is a very strongly shared belief that regardless of the backgrounds of these children, and the difficulties and barriers they face, actually, we can have success and they can make progress and they can succeed. We want to inspire children to be lifelong learners. But, we also want to model that lifelong learning, and the teachers are consistently learning alongside the pupils' (T21HP).*

In contrast to schools where there was a culture of learning and development, some teachers described school environments that reflected a more command and control management approach. In one case, Visible Classroom was reported as being used as a 'surveillance tool' that allowed senior staff to more closely monitor staff. The

impact of this was that the programme was felt to have reinforced a negative atmosphere and as a result teachers said that they did not like the intervention or find it beneficial.

> *'Sitting there with [member of SLT] and talking ... it just felt like you were doing things for the sake of it, and it just didn't sit well and the score was still low. Inside I was going, "I know that it doesn't need to be high because that would almost reflect a bad classroom if you are having to talk about behaviour"' (T10HP).*

In contrast to the intervention theory of change, which is based on the voluntary participation of teachers, at least one teacher perceived their engagement in Visible Classroom to be mandatory.

> *'I think someone on our Senior Leadership Team got an email and we were not given the choice … they told us we were going to the training' (T20LP).*

Conclusion

Key conclusions
1. Pupils in schools that received Visible Learning made, on average, one month less progress in combined KS2 maths and reading compared to children in the control schools. This is our best estimate, which has a high security rating. Variability around this best estimate means that possible impacts of the study include larger negative effects and negligible positive effects.
2. A negative impact of two months reduction in learning was estimated for KS2 reading outcomes for students in Year 5. The variability around this best estimate means that, while the possible results are consistently negative, the magnitude may be larger or smaller than a two month reduction.
3. Among children eligible for free school meals (FSM), those in the treatment schools made no additional progress in the KS2 maths and reading outcome measure compared to those in the control groups. These results may have lower security than the overall findings because of the smaller number of pupils. Variability around the estimate indicates that possible effects range between substantial positive and negative impacts.
4. The IPE found that iterative feedback was especially well received in schools with an open working culture that encouraged feedback compared to schools where feedback was emphasised less.
5. There were a number of implementation barriers that may have hindered the success of the intervention. Examples included teachers not having enough time to properly engage with the feedback they were given and, in some cases, feeling self-conscious about lessons being recorded.

Interpretation

Visible Classroom is a well-evidenced intervention based on a thorough review of relevant studies. An earlier pilot study concluded that the intervention was feasible and had the potential to improve teacher practice and therefore student attainment. This current evaluation was aimed at assessing the efficacy of the intervention under ideal conditions, which in this case involved the developers (from the University of Melbourne) providing training to participating teachers, coding lesson transcripts, and generating tailored feedback reports.

The result of the current trial is that it does not provide evidence that Visible Classroom is effective in improving student KS2 outcomes. We found no statistically significant impact of the treatment on combined KS2 maths and reading scores for Years 5 and 6 in our primary analysis, and we found evidence that the treatment had a negative effect on KS2 reading scores for Year 5 in our secondary analysis (the equivalent of two months less progress). The primary analysis results remained consistent for the subgroup of pupils eligible for FSM: among FSM pupils the intervention did not have a statistically significant impact on combined KS2 maths and reading scores.

Given that Visible Classrooms is a well-evidenced intervention, the fact that the trial indicated that it did not improve outcomes for pupils was surprising. This is compounded by the fact that the analysis also indicated that pupil outcomes were negative for those teachers who reported that they highly valued feedback, which is a very counterintuitive result. Within this context, the findings of the IPE evaluation can help explain the null result of the impact evaluation. A number of barriers to implementation and effectiveness were identified. These included feelings of self-consciousness among teachers about being recorded and assessed, which may have negatively impacted their quality of teaching. In addition, the lack of clarity around what to do when the class is interrupted, issues recording lessons, and other instructions may have taken time away from teaching and impacted teachers' ability to provide the highest quality of teaching. Finally, the pressurised nature of school life, including the need to focus on preparing pupils for SATs, could have undermined teachers' ability to make best use of the programme.

In addition to issues around implementation, school culture was identified as an important moderator of the intervention. In schools where there was a positive culture of learning and development, the intervention was

seen as having synergy with existing practice. In contrast, in schools that were described as lacking an open, supportive approach to learning and development, the intervention was experienced as less constructive. Perhaps this could have been addressed by reinforcing with school management what the aims of the programme are and how it should be presented to more junior teaching staff who are using the intervention. Nevertheless, the net result of this may have been that the intervention did not lead to a major improvement in schools that were already attuned to its basic philosophy and was not powerful enough to engender a change in culture in schools with an antithetical approach to learning and development.

While there were a range of barriers that could have negatively affected the impact of the intervention, it is important to note that there was also evidence that supported its approach. Participants described the feedback as being clear and its 'objective' and 'evidence-based' character was felt to be more accurate than other feedback experienced by some teachers as well as being easier to apply. In addition, there were examples of teachers who had clearly understood the value of the iterative learning cycle and changed mindset that are core elements of Visible Classroom. One interpretation of this, therefore, is that the net effect of the barriers and facilitators detailed above was that within this context the intervention is not effective. An alternative interpretation is that the failure of the programme to lead to improvements is not due to the presence of barriers, but that the assumptions underlying the programme are not borne out in reality. This interpretation is supported by the fact that there was a negative association between teachers valuing feedback and improved pupil outcomes (in addition to a negative but non-significant association between compliance and pupil outcomes). In either case, the evidence suggests that the use of Visible Classroom in schools should be reviewed and that further development and testing should take place before it is recommended as an approach for improving teaching quality and pupil outcomes.

Limitations

Given the strong availability of KS1 and KS2 data from the NPD, the trial did not face high levels of attrition at the student level. However, the difficulty of school recruitment proved to be a limiting factor in the trial. The reduced number of schools compromised the evaluation, increasing the minimum detectable effect size, as did changing the scope of the evaluation to explore the effects of a longer, two-year intervention on student learning outcomes. This also meant that a number of Year 5 students who were taught by a VC teacher during their Year 6 were effectively exposed to treatment for an additional year relative to rest of the study sample. We were unable to determine for whom this was the case and thus unable to control for this in the analysis. It is important to note that most of the confidence intervals in the specifications included zeros and positive values meaning that a more powered trial might have confirmed the negative impacts of specifications that are not significant, or potentially found small positive impacts. Nevertheless, the overall conclusion of the evaluation, which is that the intervention appeared to have no significant impact, would be unlikely to change.

Additionally, it is worth considering the limitations of the scope of the IPE evaluation. The IPE explored acceptability of the intervention amongst teachers and schools; however, it did not consider student responses to the intervention's introduction into their classrooms. Novel technology in schools, while helpful, could potentially become a distraction. Children may be more focused on the new piece of technology than on the bigger picture of allowing the intervention take place properly. In addition, the IPE does not explore in great depth the activity and perspectives of teachers in the control schools. This approach was chosen, in part, because the IPE was trying to understand implementation practices in schools where the intervention had taken place. Exploring this, as well as other factors like years of teaching practice, teaching style, and school ethos in future work could provide further clarity on the outcome of this evaluation.

Future research and publications

As previously mentioned, future trials might include a larger sample of schools to improve the power of the evaluation. Given the degree of positive feedback from teachers in schools where an open learning environment is encouraged, perhaps exploring the intervention in such contexts could help ascertain whether the overall findings of this evaluation are consistent when restricted to schools that operate in a working culture that promotes openness and feedback.

It is not the intention of the evaluation team to seek to publish the findings from the evaluation.

References

Alberts, H. J., Martijn, C. and De Vries, N. K. (2011) 'Fighting Self-Control Failure: Overcoming Ego Depletion by Increasing Self-Awareness', *Journal of Experimental Social Psychology*, 47 (1), pp. 58–62.

Allen, J., Gregory, A., Mikami, A., Lun, J., Hamre, B. and Pianta, R. (2013) 'Observations of Effective Teacher-Student Interactions in Secondary School Classrooms: Predicting Student Achievement with the Classroom Assessment Scoring System—Secondary', *School Psychology Review*, 42 (1), pp. 76–98.

Braun, V. and Clarke, V. (2006) 'Using Thematic Analysis in Psychology', *Qualitative Research in Psychology*, 3 (2), pp. 77–101.

Arnold, I. (2011) 'John Hattie: Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement' (book review), *International Review of Education*, pp. 57, 219.

Gerber, A. S. and Green, D. P. (2012) *Field Experiments: Design, Analysis and Interpretation*, New York: WW Norton.

Rowe, K. (2003) 'The Importance of Teacher Quality as a Key Determinant of Students' Experiences and Outcomes of Schooling': https://research.acer.edu.au/research_conference_2003/3

Hattie, J. (2008) *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*, Abingdon: Routledge.

Hedges, L. and Hedberg, E. (2007) 'Intraclass Correlation Values for Planning Group-Randomized Trials in Education', *Educational Evaluation and Policy Analysis*, 29 (1), pp. 60–87.

Hoffman, T. et al. (2014) 'Better Reporting of Interventions: Template for Intervention Description and Replication (TIDieR) Checklist and Guide', *BMJ*, p. 348.

Pintrich, P. R. and Schunk, D. H. (2002) 'Motivation in Education', *Creative Education*, 5 (14).

Robinson, L., Adair, P., Coffey, M., Harris, R. and Burnside, G. (2016) 'Identifying the Participant Characteristics That Predict Recruitment and Retention of Participants to Randomised Controlled Trials Involving Children: A Systematic Review', *Trials*, 17 (1), p. 294.

Skipp, A. and Tanner, E. (2015) 'The Visible Classroom: Evaluation Report and Executive Summary', London: Education Endowment Foundation. https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/EEF_Project_report_TheVisibleClassroom.pdf

Appendix A: EEF cost rating

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

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

Appendix B: Security classification of trial findings

OUTCOME: *KS2 maths and reading for Y5*

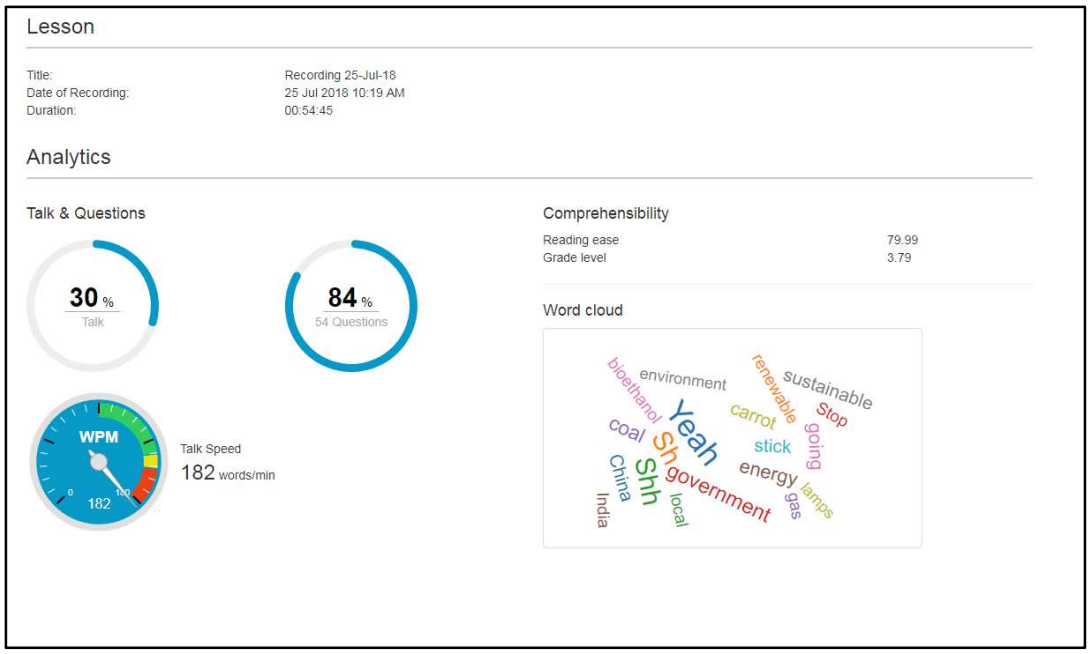
Please use this template to assign a separate security rating for each primary outcome.

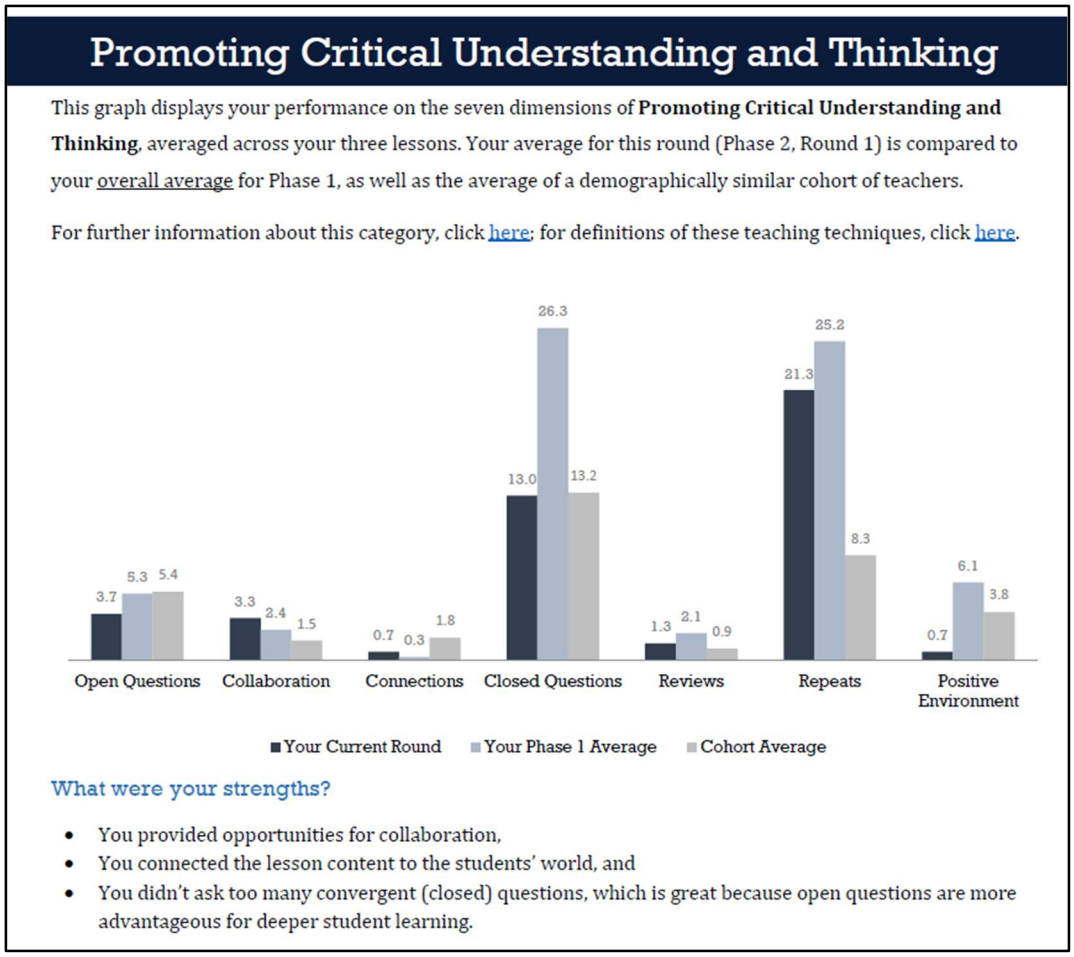
Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	**MDES**	**Attrition**					
5	Randomised design	<= 0.2	0-10%					
4	Design for comparison that considers some type of selection on unobservable characteristics (e.g. RDD, Diff-in-Diffs, Matched Diff-in-Diffs)	0.21 - 0.29	11-20%	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

Threats to validity	Threat to internal validity?	Comments
Threat 1: Confounding	Low	Randomisation was conducted by the independent evaluation team and code published in the report. Capping of teachers was included as a stratum in randomisation to avoid bias. Imbalance of 0.05 at baseline that was controlled for.
Threat 2: Concurrent Interventions	Low	There is no evidence of similar interventions in the control arm although it is likely that some may have accessed alternative teacher training and feedback programmes, which would reflect 'business as usual'.
Threat 3: Experimental effects	Moderate	The evaluators do not explicitly report teacher practice changes in the control group. Further detail on this teacher practice data would reduce risk of this threat.
Threat 4: Implementation fidelity	Moderate/ High	Compliance was moderate (80% of pupils had a compliant teacher). The definition of compliance was relaxed during the trial and the programme was not implemented fully as intended with some receiving a two-year programme and some a one-year, more intensive programme.
Threat 5: Missing Data	Low	Only 5% missing data and not differential between arms.
Threat 6: Measurement of Outcomes	Low	KS2 data from national high stakes tests, blind to allocation and no floor/ceiling effect.
Threat 7: Selective reporting	Low	Trial registered and analysis follows SAP, with some small changes from protocol. All planned outcomes are reported.

- **Initial padlock score:** 4 Padlocks – a large, well-run trial with MDES of 0.18 at randomisation and attrition of 5%.
- **Reason for adjustment for threats to validity**: 1 Padlock – Lack of information on concurrent interventions or experimental effects and some issues with implementation.
- **Final padlock score:** initial score adjusted for threats to validity = 4 Padlocks

Appendix C: Example of teacher feedback report

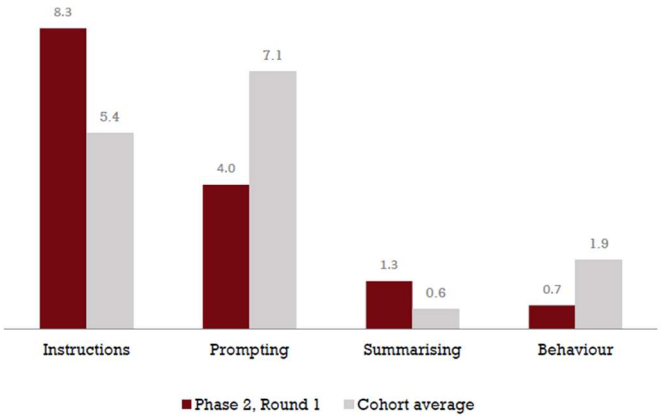




Promoting Engagement and Understanding

This graph displays your performance on the four dimensions of **Promoting Engagement and Understanding**, averaged across your three lessons. Your average for this round (Round 1) is compared to the average of a demographically similar cohort of teachers.

For further information about this category, click here; for definitions of the teaching techniques, click here.



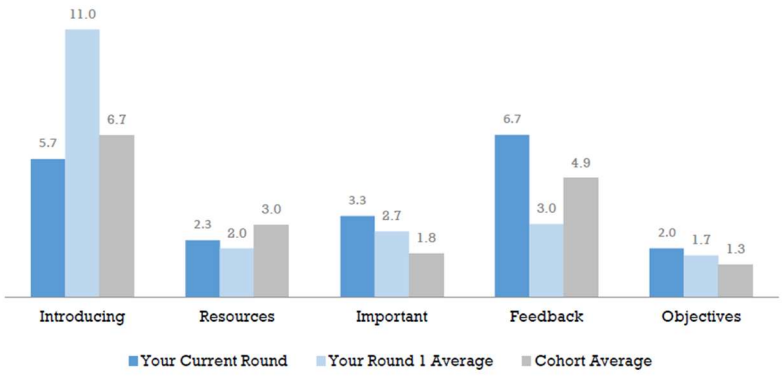
What were your strengths?

- You provided clear, step-by-step instructions, and
- You summarised lesson content frequently.

Providing Feedback and Additional Instruction

This graph displays your performance on the five dimensions of **Providing Feedback and Additional Information**, averaged across your lessons. Your average for this round (Phase 2, Round 2) is compared to your average for Phase 2, Round 1, as well as the average of a demographically similar cohort of teachers.

For further information about this category, click here; for definitions of the teaching techniques, click here.



What were your strengths?

- You emphasised important points,
- You provided immediate, specific, and corrective feedback, and
- You established clear learning objectives throughout your lessons.

Appendix D: Abridged version of scoring rubric

Code	Abbreviation	Description	Examples
		Promote Critical Analysis/Thinking	
1	Deepen understanding	Provides students with opportunities to deepen understanding through sustained context-specific dialogue and open-ended or divergent questions (i.e. those for which there is no one correct answer). Critical features: dialogue; open-ended questions; goal of deepening understanding.	
2	Collaboration	Provides students with opportunities to collaborate with peers and actively engage in learning. Critical features: collaboration among peers; active learning.	
3	Connections	Helps students make specific connections between classroom content and the real world. Critical feature: real world connection.	
4	Convergent	Checks factual recall by asking closed questions (i.e. those that have yes/no answers). Or, asks questions that have only one correct answer or a set of possible correct answers. Critical feature: closed questions.	

5	Review	Reviews previously learned concepts at the beginning of the lesson or throughout the lesson. Refers back to material learned in previous lessons. Critical feature: reviews content from previous lessons. Differs to: Summarise because it refers to content covered in past lessons rather than content covered in this lesson.	
6	Repeats comment	Repeats comment or question from students before answering in order to encourage and reinforce correct answers. Critical feature: repeats student comment/question.	
7	Positive classroom environment	Creates a classroom environment in which students have an opportunity to freely and spontaneously ask/provide task-related questions or feedback that is related to the content. Critical feature: Teacher creates opportunities for student questions/feedback.	
8	Instructions	Provides step-by-step instructions about how students should complete tasks/activities. Critical feature: teacher uses instructions about content-related activities. Note for coders: only code content-oriented instructions. Omit all process oriented instructions such as: "I want you all to sit down, get your books out and paste the instructions."	

9	Prompting	Teacher uses prompting or probing (i.e. a question or statement) to elicit an appropriate student response. Critical feature: prompts; probes. Differs to: deepen understanding because the question or statement is aimed at triggering student recall or response rather than deepening their understanding.	
10	Summarise	Concludes the lesson (or segment of a lesson) by re-stating or summarising key points; provides opportunities for future follow-up or engagement. Critical feature: summarises content covered in the lesson. Differs to: Review because it refers to a summary of material covered in the current lesson, not previous ones.	
11	Behaviour	Sets clear behaviour expectations and provides behaviour reminders during the lesson where appropriate or necessary. Note for coders: Omit reprimands that do not provide clear behavioural expectations (e.g. "John, ssssshhhh!" "I have told you several times to be quiet!" "OK John, go to the Principal's office!"	

12	Introduces & explains	Introduces and explains new or complicated vocabulary and terminology. Simplifies concepts by breaking them down into smaller ideas, or elaborates on abstract concepts using concrete, developmentally appropriate examples. Critical feature: Explains new or complicated words/concepts Differs to: Connections because when examples here, they are used for the purpose of explaining a word or idea rather than taking an already learned concept and showing how it relates to the real world.	
13	Resources	Uses written/visual/audio resources to support learning. Critical feature: refers to additional resource.	
14	Important	Teacher emphasises important points. The teacher may not use the word 'important' but the tone and context of the dialogue suggest that this information should be prioritised over others. Critical features: highlights that particular information is important and should be paid attention to. Differs to: Summarise because the teacher is not providing a review of what has been talked about earlier but rather highlighting specific information as being of particular importance. The teacher is indicating a hierarchy of knowledge, with this piece of information at the top.	
15	Feedback	Teacher provides immediate, specific and corrective feedback to individual students/student groups. Critical features: provides feedback.	

16	Goals/success criteria	Teacher identifies the purpose of the learning activity (i.e. the goals and learning outcomes) at the beginning of the lesson/activity and refers to them throughout the lesson. Critical features: identifies learning goals; success criteria	

Appendix E: Copy of School Participation Agreement

School Agreement to participate in the Visible classroom project with the SSAT, the University of Melbourne and Ai Media

*Please complete and sign **two copies**, retaining one and returning the second copy (with a signature) by email to primary@ssatuk.co.uk or by post to Visible Classroom project, SSAT, 5th floor, 142 Central Street, London EC1V 8AR by **Friday 22nd July 2016.***

School Name: _____

School Postcode: _____ *Headteacher Name:* _____

Aims of the Project
This visible classroom project aims to improve pupil outcomes by supporting teachers' professional practice development. It seeks to encourage teachers to reflect on their teaching and develop classroom practice using real-time, credible evidence. The project was initially developed by researchers at the University of Melbourne (UoM), working in partnership with Ai-Media, a captioning company. This evaluation is funded by the Education Endowment Foundation (EEF), and will be independently evaluated by the Behavioural Insights Team (BIT).

Trial objective
The trial's objective is to test whether providing feedback to teachers on their audio recorded lessons has an impact on student outcomes for KS2 results in maths and English.

The main question the evaluation seeks to answer:
- Does the Visible Classroom intervention increase the educational attainment of Y5 and Y6 students (age 9-11) in combined KS2 maths and English scores?

Secondary research questions consider impact on:
- The maths and English results for each year group independently
- The maths and English results of Free School Meal (FSM) students
- The impact of students whose teachers score in the bottom third of the VC rubric, as measured at baseline

Parental agreement
Shortlisted schools are required to inform all parents about the project and to collect any Opt-Out Forms returned by parents. A template will be provided for this.

Schools will need to provide the following school and pupil level information, ensuring that any opt-out data is not included.
- School's Unique Reference Number (URN) and LAESTAB number.
- KS1 results
- Unique Pupil Numbers (UPNs) for all year 5 and 6 children (barring those whose parents opt-out).

The evaluation team will use this information to access the National Pupil Database.

The Evaluation
The evaluation is being conducted by BIT and a Randomised Control Trial (RCT) approach is being applied. Schools who agree to take part will be randomly allocated to either:
1) **An intervention group**: with Visible Classroom support, or
2) **A control group:** no intervention

Random allocation is essential to the evaluation. It is important that schools understand that they may be randomly allocated to either group and are agreeing to participate regardless of the assignment.

Use of Data
All data, including pupils' KS2 results and any other pupil data, will be treated with the strictest confidence. Pupil data will be matched with the National Pupil Database by BIT and shared with

the Department for Education, the EEF and their data contractor FFT Education, and in an anonymised form to the UK Data Archive.

SSAT will also collect data from schools that will be passed on to BIT in order to complete the random assignment of schools to intervention and control groups, and assist in process evaluation (understanding how the intervention worked).

Ai Media will be providing the captioning service which powers the Visible Classroom app. These transcripts will also be accessed by UoM staff to give feedback to participating teachers in the intervention group. Usage data from the Visible Classroom app will also be collected. Both will be made available to BIT for process evaluation purposes.

Any transfer of individual-level data between parties will be performed in a secure manner (i.e. using password encrypted files).

No individual school, teacher or pupil will be identified in any report arising from the research. No other parties will have access to the data collected.

Responsibilities

SSAT will:

- Shortlist successful schools through application and interview process.
- Provide the evaluators with a list of schools to be allocated to the control and trial groups.
- Let schools know which group they have been allocated to, either control or intervention, based on the information provided by the evaluators.
- Be the first point of contact for non-technical queries (app related queries will be handled by UoM and Ai Media staff)
- Hold a celebration event in 2018 with John Hattie for all participating schools (both control and treatment). Invitation will be contingent upon school participation throughout the trial.

Also:

For the control schools;

- Will share the evaluation findings and final case studies with the school.

For the intervention schools;

- Hold a launch event with John Hattie in November 2016.
- Provide on-going implementation support to the school.
- Collate case studies for and produce publications at the end of each year.
- Provide programme and project management for the duration of the project

BIT will:

- Conduct the random allocation of schools to trial arms.
- Store all data safely and securely.
- Analyse data from the project in order to produce impact estimates.
- Conduct the process evaluation, including analysis and reporting from this.
- Produce end of project evaluation report.

In order to participate in the project, all schools will need to:

- Consent to random allocation and commit to trial participation regardless of the trial group they are assigned to.
- Send all year 5 and 6 parents the consent and opt out form.
- Provide the evaluators with schools and pupil level data as outlined above
- Carry out baseline recording of two hours of teaching prior to the project starting
- Have commitment from all year 5 and 6 class teachers to regularly use the Visible Classroom App (3 x 5 hours of instruction) and meet with their mentor to review the feedback reports if allocated to the intervention group.
- Have year 5 and 6 teachers complete a survey and agree to complete another prior to the conclusion of the trial (Control group - online, intervention group - hard copy).

- Agree to allow teachers to record in class using an iPhone/smart phone to ensure the recording is of sufficient quality to enable transcription and coding.
- Agree to notify SSAT and BIT in the event a year 5 or 6 teacher ceases employment with the school or steps away from teaching these year levels.
- Have nominated and gained commitment from mentor teachers (members of the SMT) to participate in online webinar training, fill out a short online survey and to meet with their mentees throughout the year to discuss each detailed feedback report if allocated to the intervention group.

Additionally, schools agree to the following ***if*** assigned to the intervention group:

- Attend a training event with Professor John Hattie in November for all year 5 and year 6 class teachers to learn about the technology (1st or 7th November TBC)
- Complete two hours captioning on Visible Classroom App at the end of the project
- Complete an evaluation interview: 6-8 intervention schools will be asked to complete interviews with the evaluation team about the project. Interviews will be approximately 30 minutes
- Contribute to case studies about the project, through interviews or email feedback
- Inform SSAT if there is a change of lead teacher for the project or Headteacher at the school
- To sign up to an online group where information and updates from the University of Melbourne will be provided (KHub)
- Allow extra staff time if necessary and within reason to the project

As a school I commit to remaining a part of the visible classroom project as detailed above for the period of September 2016 – July 2018

Head teacher name: _____

Head teacher signature: _____

School name: _____

Local Authority:_____

Head teacher Signature: _____ Date: _____

Head teacher Email address: _____

School Contact (if different from head teacher): _____

School Contact email address: _____

School Tel no: _____

Thank you for agreeing to take part in this research. Please return this form to:
Email: primary@ssatuk.co.uk Or via post to: *Visible Classroom Project, SSAT, 5th floor, 142 Central Street, London EC1V 8AR*

Appendix F: Teacher Information Sheet

Visible Classroom Efficacy Trial 2016-2018

Plain Language Statement – Teacher

Dear Teacher,

My name is Associate Professor Janet Clinton and I work at the University of Melbourne in Australia. Your school has applied to take part in an exciting national research project, Visible Classroom, which aims to improve student achievement outcomes by supporting teachers' professional practice development. In order to learn how effective the Visible Classroom project is, I, along with an experienced research team at the University of Melbourne and our partners Ai Media and SSAT, will be conducting a trial to measure the efficacy of this project. We have been given approval to conduct this trial by the Behavioural Insights Team External Ethical Advisory Panel.

What is involved

The central feature of the Visible Classroom is the conversion of teacher talk into text form. Teachers will wear a microphone which records their speech generating an audio file on their mobile phone. After recording their lessons, each teacher will receive personalised feedback on their teaching. This feedback is designed to encourage teachers to reflect on their teaching and develop their classroom practice. Similar studies have found that when teachers reflect on their teaching – students' learning can improve significantly.

Your school has applied to be one of the 140 primary schools in England who will be involved in the project. Anonymous pupil data in all 140 schools will be analysed in 2018/19 to evaluate whether the Visible Classroom had an impact on test scores. The Behavioural Insights Team has been appointed to evaluate the project externally.

All participants of this project will be fully informed of what their role entails. As a teacher, you will also receive a similar letter to this one which you will be able to read to your students in order to inform them about the research project you are about to participate in.

Confidentiality

We intend to protect your anonymity and the confidentiality of your information to the fullest possible extent, within the limits of the law. Your data will be de-identified by creating a unique identification number. Your name and contact details will be kept in a separate, password-protected computer folder. This will only be able to be linked to your data by the researchers, for example, in order to check accuracy of data. All data presented in the final reports will be a combination of many individual responses and you will in no way be identified personally. Quotes used to demonstrate the captioning technology will be de-identified and pseudonyms will be used. The use of this data is for research purposes only and will not have any impact on your relationship with your school.

Once this research has been completed, a brief summary of the findings will be available to you. Results will be presented at academic conferences and published in academic journals. The data will be kept securely in the Melbourne Graduate School of Education for five years from the date of publication, before being destroyed.

Because there are several different members of the school participating in this project, including teachers, students and leaders, it is important to us that issues are addressed immediately and responsibly. In the unlikely event that a teacher or student tells us any important things that may concern their health or safety we will inform both the parents of any students involved and the school.

Consent

Should you wish to withdraw from the project at any stage, or to withdraw any unprocessed data you may have supplied, you are free to do so without prejudice. To withdraw your data from this research, please send an email to sarah.mason@unimelb.edu.au, with your full name, date of birth, and your contact email.

Should you require any further information, or have any concerns, please do not hesitate to contact the Project Manager, Research Fellow, Sarah Mason sarah.mason@unimelb.edu.au Should you have any concerns

about the conduct of the project, you are welcome to contact the Manager, Human Research Ethics, The University of Melbourne, on ph: +61 03 8344 2073, or fax: +61 03 9347 6739.

In the unlikely event, that the research project makes anyone feel uncomfortable they can let us know and we will refer them to someone who can talk to them about their feelings.

Yours sincerely,

Associate Professor Janet Clinton

Appendix G: Parent Information Sheet and Opt-Out Form

Visible Classroom Efficacy Trial 2016-2018

Plain Language Statement – Parent/Carer

Dear Parent/Carer,

My name is Associate Professor Janet Clinton and I work at the University of Melbourne in Australia. Your child's school has applied to take part in an exciting national research project, Visible Classroom, which aims to improve student achievement outcomes by supporting teachers' professional practice development. In order to learn how effective the Visible Classroom project is, I, along with an experienced research team at the University of Melbourne and our partners Ai Media and SSAT, will be conducting a trial to measure the efficacy of this project. We have been given approval to conduct this trial by the Behavioural Insights Team External Ethical Advisory Panel.

What is involved

The central feature of the Visible Classroom is the conversion of teacher talk into text form. Your child's classroom teacher will wear a microphone which records their speech generating an audio file on their mobile phone. After recording their lessons, each teacher will receive personalised feedback on their teaching. This feedback is designed to encourage teachers to reflect on their teaching and develop their classroom practice. Similar studies have found that when teachers reflect on their teaching – students' learning can improve significantly.

Your child's school has applied to be one of the 140 primary schools in England who will be involved in the project. Anonymous pupil data in all 140 schools will be analysed in 2018/19 to evaluate whether the Visible Classroom had an impact on test scores. The Behavioural Insights Team has been appointed to evaluate the project externally.

For the purpose of research, information provided by your child's school (unique pupil number) will be linked with information about your child from the National Pupil Database (held by the Department for Education) and shared with the evaluators, the Department of Education, and in an anonymous form to the UK Data Archive.

We may ask your child to take part in a survey or focus group; again all data will be treated with strictest confidence. A separate letter and consent form will be sent to you regarding this part of the research.

Your child's teacher will receive a similar letter to this one which they will be able to read to your child to inform them about the research project.

Confidentiality

We intend to protect your child's anonymity and the confidentiality of their responses to the fullest possible extent, within the limits of the law. Your child's data will be de-identified by creating a unique identification number. Their name and contact details will be kept in a separate, password-protected computer folder. This will only be able to be linked to their data by the researchers, for example, in order to check accuracy of data. All data presented in the final reports will be a combination of many individual responses and your child will in no way be identified personally. Quotes used to demonstrate the captioning technology will be de-identified and pseudonyms will be used. The use of this data is for research purposes only and will not have any impact on your child's relationship with their school or their general education.

Once this research has been completed, a brief summary of the findings will be available to you. Results will be presented at academic conferences and published in academic journals. The data will be kept securely in the Melbourne Graduate School of Education for five years from the date of publication, before being destroyed.

In the unlikely event that a participant tells us any important things that may concern their health or safety we will inform both the parents and the school.

Consent

Please be advised that if you **DO NOT** want your child to participate you must sign and return the opt-out form (attached), by not opting out, you are voluntarily consenting for your child to participate in this research. Should

you or your child wish to withdraw at any stage, or to withdraw any unprocessed data you or your child may have supplied, you are free to do so without prejudice. To withdraw your child's data from this research, please send an email to sarah.mason@unimelb.edu.au, with your child's full name, date of birth, and your contact email.

Should you require any further information, or have any concerns, please do not hesitate to contact the Project Manager, Research Fellow, Sarah Mason sarah.mason@unimelb.edu.au Should you have any concerns about the conduct of the project, you are welcome to contact the Manager, Human Research Ethics, The University of Melbourne, on ph: +61 03 8344 2073, or fax: +61 03 9347 6739.

In the unlikely event, that the research project makes anyone feel uncomfortable they can let us know and we will refer them to someone who can talk to them about their feelings.

Yours sincerely,

Associate Professor Janet Clinton

Visible Classroom Efficacy Trial 2016-18

Opt-Out Form

If you **DO NOT** want information about your child to be shared for use in the Visible Classroom Efficacy Trial 2016-18, please **return this form to your child's school** by [INSERT DATE].

I **DO NOT** want information about my child to be shared for use in the *Visible Classroom project* evaluation

Parent Name: ……..

Parent/Carer Signature: ……..…….

Date: ……

Child's Name: ……..

Child's Date of Birth: ……..

Child's School: …….

Appendix H: Interview sample characteristics

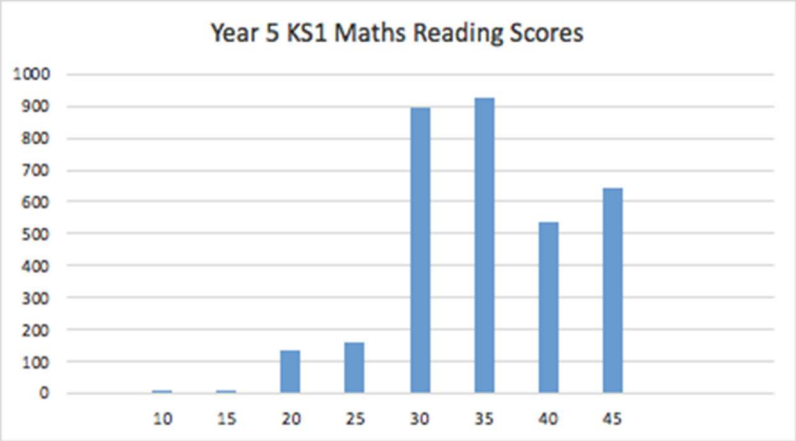
Alias	Y1 or Y2	Engagement	Ofsted	Role
T1MP[18]	1	Medium	Good	Teacher
T2HPM	1	High	Good	Teacher /Mentor
T3HP	1	High	Good	Teacher
T4HP	1	High	Good	Teacher
SLT1M	1	SLT	Good	Mentor
T5HP	1	High	Good	Teacher
T6HP	1	High	Good	Teacher
SLT2M	1	SLT	Good	Mentor
T7HP	1	High	RI	Teacher
T8HP	1	High	RI	Teacher
T9HP	1	High	RI	Teacher
T10HP	1	High	RI	Teacher
T11MP	1	Medium	RI	Teacher
T12HP	1	High	RI	Teacher
T13HP	1	High	RI	Teacher
SLT3M	1	SLT	RI	Mentor
SLT4M	1	SLT	RI	Mentor
SLT5M	1	SLT	RI	Mentor
SLT6M	1	SLT	Good	Mentor
T14HP	1	High	Good	Teacher
T15MP	2	Medium	Good	Teacher

[18] These identifiers help the reader to know pertinent information about the teacher, without removing anonymity. For example, this means Teacher 1 from a school with medium engagement who is a participant in the intervention (rather than mentor).

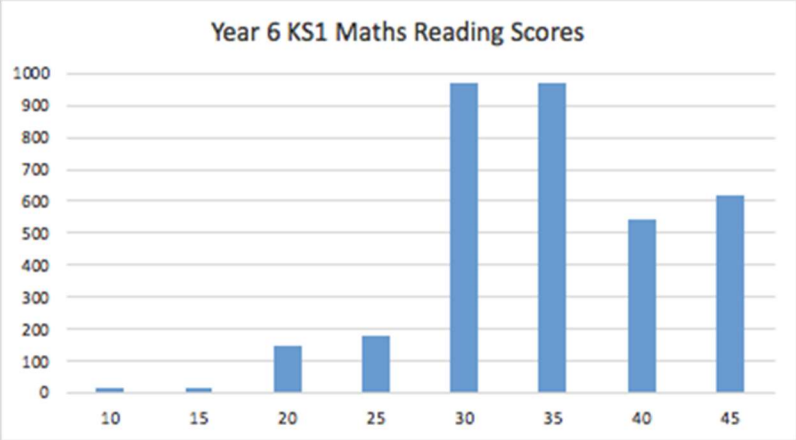
T16HP	2	High	Good	Teacher
T18HP	2	High	Outstanding	Teacher
T19LPM	2	Low	Good	Teacher /Mentor
T20LP	2	Low	Good	Teacher
T21HP	2	High	Good	Teacher /Mentor

Appendix I: Histograms of pre-test scores

Figure A1: KS1 Maths and Reading scores for Year 5



N = 3318. Please note: we have omitted scores above 50 in accordance with ONS rules to exclude any bins with frequencies <10.



N = 3481. Please note: we have omitted scores above 50 in accordance with ONS rules to exclude any bins with frequencies <10.

Appendix J: Code

Code H.1: Randomisation

```
clear

set more off

cd "C:\Users\daniel.carr\Google Drive\EEF\Visible Classroom Evaluation Project"

*import data

************************************************************


*VC schools - this is the list of schools in the trial

import  excel "EEF VC schools for randomisation V2", sheet("Sheet1") first

rename _all, lower

*drop if no spreadsheet returned (these are not to be randomised).

drop if spreadsheetreturned==""

*note - sorting URN for all data files pre-merge as School Unique Reference Numbers
will be used for matching in our merge.

sort urn

save "VC list2.dta", replace

*note: the 'capped' variable is included in this file and is coded as a yes or
no.

clear

*KS1 data - this is a data file from DfE's School Performance Comparison website
for 2014-15 KS2 students, that also contains there KS1 results in Average Point
Score (tks1aps) form.

import delimited "england_ks2"

rename _all, lower

keep urn schname tks1aps

drop if urn==.

destring tks1aps, replace

sort urn

save "DfE ks2.dta", replace

clear

*FSM data - this is also a data file from the above source, but here uses the
school-wide census, from which we draw the total proportion of 'FSM ever' students
(those who've been on FSM, or are currently on FSM)
```

```
import delimited "england_census"

rename _all, lower

keep urn pnumfsmever


drop if urn=="NAT"

destring urn, replace

drop if urn==.

*put FSMever into decial form

destring pnumfsmever, generate(percentfsm) ignore(`"%"') force percent

sort urn

save "DfE census.dta", replace

***** merge - original school list with the two DfE data sources to get our FSM
and KS1 APS variables for the stratification.

clear

use "VC list2.dta"

merge 1:1 urn using "DfE ks2", gen(_ks2merge)

keep if _ks2merge==3

*all schools matched, but three have no information on KS1 APS.

merge 1:1 urn using "DfE census", gen(_censusmerge)

drop if _censusmerge==2

*note: 4 schools not matched with census data - URNs not present in the census
spreadsheet (oddly FSM data is in the KS2 spreadsheet for the 2014-15 cohort of
year 6 students though). Will use the 'missing' command in bitrandomise to create
a strata for these schools.

**** create categorical FSM and KS1 variables that split schools into high and
low.

*FSM

destring percentfsm, replace

egen fsmmed = median(percentfsm)

generate fsm = 1 if percentfsm > fsmmed

replace fsm = 0 if percentfsm <= fsmmed

replace fsm = . if percentfsm ==.

label def fsm 1 "High share" 0 "Low share"

label val fsm fsm
```

```
*KS1

destring tks1aps, replace

egen ks1med = median(tks1aps)

generate ks1 = 1 if tks1aps > ks1med

replace ks1 = 0 if tks1aps <= ks1med

replace ks1 = . if tks1aps ==.

label def ks1 1 "High" 0 "Low"

label val ks1 ks1

save "VC merged.dta", replace

**** randomise

*include the three stratifying variables, and use the missing command. Ratio is
1:1.

bitrandomise capped fsm ks1, gen(allocation) seed(1454331) groups(1 1) missing

*missing data causing an issue? 3 schools have no KS1 data, and 4 (different)
schools no FSM data.

label def allocation 1 "Control" 2 "Treatment"

label val allocation allocation

drop ks1 capped spreadsheetreturned schname tks1aps _ks2merge pnumfsmever
percentfsm _censusmerge fsmmed fsm ks1med

sort allocation

export excel using "EEF VC trial arm allocations", sheetreplace
firstrow(variables)

save "VC allocated.dta", replace
```

Code H.2: Analysis

```
vers 14
set more off

cap cd "C:\Users\Hal The Destroyer\Desktop\EEF - URLEY and VC\Visible Classroom"
cap  cd "P:\Working"


*use "dta_clean\Visibleclean_forONS.dta", clear

//appending 2017 and 2018 ONS data
import delimited P:\Working\KS2Pupil_2017_KS1_Census_v2.txt, varnames(1) clear
tostring _all, replace
tempfile 2017
sa `2017', replace

import delimited P:\Working\KS2Pupil_2018_KS1_Census_v2.txt, varnames(1) clear
tostring _all, replace
tempfile 2018
sa `2018', replace
append using `2017'

//fill year of birth for the two school years
       gen     yob=yearofbirth_spr18
       replace yob=yearofbirth_spr17 if yearofbirth_spr18==""
```

```
        destring yob, replace

        gen     mob=monthofbirth_spr18
        replace mob=monthofbirth_spr17 if monthofbirth_spr18==""
        destring mob, replac

//fill in everfsm indicator based on appropriate school year
gen     everfsmspryr=everfsm_6_p_spr17 if year_num=="5"
replace everfsmspryr=everfsm_6_p_spr16 if year_num=="6"

//Need age in months -- pick date during intervention
gen birth=ym(yob, mob)
format birth %tm
gen intervention=ym(2018, 9)
format intervention %tm
gen agem=intervention-birth
gen age_years=agem/12

*****************************************************
*******************Primary Analysis******************
*****************************************************
//Combined KS2 Maths and Reading results, analysed separately by year level
//Regress outcome on treatment, strata (FSM students split by above/below sample median, average
point score split similarly, schools offered entry capped or uncapped), student-level KS2 baseline
//Standard errors clustered at school level

// [] Construct KS1 baseline variable from math and reading components -- sensitivity analysis if
missing >5% outcome data

// [] School-level numeric identifiers (for clustering) --  just use school_urn

// Add notes
#delimit ;
loc tablenote "Standard errors clustered at the school level.
+ p < 0.10, * p < 0.05, ** p < 0.01"
;
#delimit cr

//destringing and encoding
global cont "ks2_matmrk ks2_readmrk ks2_ks1matps ks2_ks1readps year_num"
global cat  "fsm ks1 capped allocation school_urn ks2_gender everfsmspryr"
foreach y of global cont {
        replace `y'="" if `y'=="."
        destring `y', replace
}

gen fsm_num = .
replace fsm_num = 1 if fsm=="High share"
replace fsm_num = 0 if fsm=="Low share"
label define fsmlab 0 "Low share" 1 "High share"
label val fsm_num fsmlab
tab fsm fsm_num, m //missing for schools 108006, 116485, 125027. missings balanced
drop fsm
rename fsm_num fsm

gen ks1_num = .
replace ks1_num = 1 if ks1=="High"
replace ks1_num = 0 if ks1=="Low"
label define ks1lab 0 "Low" 1 "High"
label val ks1_num ks1lab
tab ks1 ks1_num, m //missing for schools 138257, 138668. all missings are in treatment
drop ks1
rename ks1_num ks1

gen capped_num = .
replace capped_num = 0 if capped=="No"
replace capped_num = 1 if capped=="Yes"
label define yesno 0 "No" 1 "Yes"
label val capped_num yesno
tab capped capped_num, m
drop capped
rename capped_num capped

gen allocation_num = .
replace allocation_num=1 if allocation=="Treatment"
replace allocation_num=0 if allocation=="Control"
label define alloclab 0 "Control" 1 "Treatment"
label val allocation_num alloclab
tab allocation allocation_num
drop allocation
rename allocation_num allocation
```

```
/*
replace school_urn="" if school_urn=="."
encode school_urn, gen(school_urni)
drop school_urn
rename school_urni school_urn
*/
destring school_urn, replace

gen ks2_gender_num = .
replace ks2_gender_num = 0 if ks2_gender=="M"
replace ks2_gender_num = 1 if ks2_gender=="F"
label define glab 0 "Male" 1 "Female"
label val ks2_gender_num glab
label var ks2_gender_num "Female=1"
tab ks2_gender ks2_gender_num,m
drop ks2_gender
rename ks2_gender_num female

gen everfsm = .
replace everfsm = 0 if everfsmspryr == "0"
replace everfsm = 1 if everfsmspryr == "1"
tab everfsmspryr everfsm // missing for 52 students
drop everfsmspryr
rename everfsm everfsmspryr



//Primary outcome measure: combined raw KS2 Maths and raw KS2 Reading
gen       ks2_mathsread = ks2_matmrk + ks2_readmrk
label var ks2_mathsread "Combined KS2 Maths and Reading - primary outcome"

gen       ks1_mathsread = ks2_ks1matps + ks2_ks1readps
label var ks1_mathsread "Combined KS1 Maths and Reading - baseline control"

eststo clear

forvalues i = 5/6 {
       loneway ks1_mathsread school_urn if year_num==`i' // for ICC baseline
       loneway ks2_mathsread school_urn if year_num==`i' //for ICC endline
}

       loneway ks1_mathsread school_urn  // for ICC baseline overall
       loneway ks2_mathsread school_urn  // for ICC baseline overall

//Create strata dummies
gen     block = 1 if fsm==0 & ks1==0 & capped==0
replace block = 2 if fsm==0 & ks1==0 & capped==1
replace block = 3 if fsm==0 & ks1==1 & capped==0
replace block = 4 if fsm==0 & ks1==1 & capped==1
replace block = 5 if fsm==1 & ks1==0 & capped==0
replace block = 6 if fsm==1 & ks1==0 & capped==1
replace block = 7 if fsm==1 & ks1==1 & capped==0
replace block = 8 if fsm==1 & ks1==1 & capped==1

ta block, ge(dblock)

//provisional
gen dblock9 = 0 // this is the missing strata bin that we've created so that we don't lose these
peeps for analysis

replace dblock9 = 1 if missing(dblock1)
forvalues c = 1/8 {
       replace dblock`c' = 0 if missing(dblock`c')
       }

replace block = 9 if dblock9==1

forvalues b = 1/9 {

       codebook school_urn if block==`b' & year_num==5
       codebook school_urn if block==`b' & year_num==6


       }

*****************************************************
********************Missing Data*********************
*****************************************************
```

```
*Asfordby Captain's and William Davie schools are missing DOB, so they are dropped from balance
checks
/*
ONS - need to check degree to which KS1 and KS2 data are missing. Hopefully it's all <5% and we can
ignore sensitivity analysis!
If >5% of outcome data are missing, we will fit a regression where the outcome is the missingness of
Y and where explanatory variables are student level prior attainment, gender, and FSM, Ofsted rating
of school

If missing data are systematic, we will conduct sensitivity analysis. Impute missing scores for
pupils using multiple imputation, including KS1 attainment, census (FSM status) and school-level
data (school-level perfomrnace in the baseline data)

Additional sensisitivty analysis for pupils with KS2 attainment data but not KS1 attainment data.
KS1 performance will be imputed using the school's mean KS1 score.

Pupils missing KS1 and KS2 attainment data will be discarded from analysis.

Multiple imputation would be required if >5% of the outcome data is missing
*/

//
gen m_ks2 = mi(ks2_mathsread)
tab m_ks2              // this must be <= 5% =>2.4%
tab m_ks2 if year_num==5             // this must be <= 5% => 2.8%
tab m_ks2 if year_num==6             // this must be <= 5% => 2.0%

//in both regressions those with lower ks1 attainment are most likely to be missing
//however crucially the intervention is not significant in either regression
reg m_ks2 allocation dblock2-dblock9 ks1_mathsread if year_num==5, cluster (school_urn)
reg m_ks2 allocation dblock2-dblock9 ks1_mathsread if year_num==6, cluster (school_urn)

/*
Missing KS1 data
*/
gen m_ks1 = mi(ks1_mathsread)
tab m_ks1 // =>4.27% missing
tab m_ks1 if year_num==5 // =>3.35%
tab m_ks1 if year_num==6 // =>5.12%

/*As a sensitivity check re-run the analysis using the mean-imputed value
for missing ks1 values*/

//drop anyone with only math or only reading for primary analysis
preserve
        drop if missing(ks2_matmrk) | missing(ks2_readmrk) //drops 170 obs out of 7,104
        //284 students with missing ks1_mathsread therefore do school level mean imputation
        gen imp_ks1_mathsread=ks1_mathsread // mean imputation code

        //year5
        forvalues i = 5/6 {
                bys school_urn: egen avgks1_mathsread_`i'=mean(ks1_mathsread) if year_num==`i'
                replace imp_ks1_mathsread=avgks1_mathsread_`i' if imp_ks1_mathsread==.
        }

        global x1  "allocation ks1_mathsread dblock2-dblock9" //as intended and specified in SAP
with strata dummies
        global xi1 "allocation imp_ks1_mathsread dblock2-dblock9" // using imputed values for
missing pre scores

/*
//MDES calculations -- bc of preserve/restore limitations cannot run this and then do rest of
analysis, so running this once to get numbers and then blocking out

                        //full analysis sample without imputed
                        forvalues i = 5/6 {
                                loneway ks1_mathsread school_urn if year_num==`i' // for ICC baseline
                                loneway ks2_mathsread school_urn if year_num==`i' //for ICC endline
                                tab allocation if year_num==`i' // student level split
                        }

                                //year 5
                                preserve
                                keep if year_num==5
                                pwcorr ks2_mathsread ks1_mathsread //pre post test correlation

                                gen present = 1
                                collapse (count) present (min) allocation capped fsm ks1,
by(school_urn) // need to get cluster sizes
                                sum present // avg cluster size for year 5:
                                tab allocation, m //school-level split
```

```
                                restore

                                //year 6
                                preserve
                                keep if year_num==6
                                pwcorr ks2_mathsread ks1_mathsread //pre post test correlation

                                gen present = 1
                                collapse (count) present (min) allocation capped fsm ks1,
by(school_urn) // need to get cluster sizes
                                sum present // avg cluster size for year 6:
                                tab allocation, m //school-level split
                                restore

                        //keeping only fsm kids

                        keep if everfsmspryr==1

                        forvalues i = 5/6 {
                                loneway ks1_mathsread school_urn if year_num==`i' // for ICC baseline
                                loneway ks2_mathsread school_urn if year_num==`i' //for ICC endline
                                tab allocation if year_num==`i' // student level split
                        }


                                //year 5
                                preserve
                                keep if year_num==5
                                pwcorr ks2_mathsread ks1_mathsread //pre post test correlation

                                gen present = 1
                                collapse (count) present (min) allocation capped fsm ks1,
by(school_urn) // need to get cluster sizes
                                sum present // avg cluster size for year 5:
                                tab allocation, m //school-level split
                                restore

                                //year 6
                                preserve
                                keep if year_num==6
                                pwcorr ks2_mathsread ks1_mathsread //pre post test correlation

                                gen present = 1
                                collapse (count) present (min) allocation capped fsm ks1,
by(school_urn) // need to get cluster sizes
                                sum present // avg cluster size for year 6:
                                tab allocation, m //school-level split
                                restore
*/

/*
        //Analysis comparison of characteristics

                //Year 5
                preserve
                keep if year_num==5
                gen present=1
                collapse (count) present (min) allocation, by(school_urn) // need to get cluster
sizes
                sum present
                merge m:1 school_urn using "P:\Working\School urns with dfe info_for ONS_v3.dta"
                drop if _merge==2

                di in red "Year 5"
                tab rating_num allocation, m
                tab establishment_num allocation, m
                tab urbanrural_num allocation, m
                sum numberofpupils if allocation==0
                sum numberofpupils if allocation==1
                restore

                //Year 6
                preserve
                keep if year_num==6
                gen present=1
                collapse (count) present (min) allocation, by(school_urn) // need to get cluster
sizes
                sum present
                merge m:1 school_urn using "P:\Working\School urns with dfe info_for ONS_v3.dta"
                drop if _merge==2
```

```
		di in red "Year 6"
		tab rating_num allocation, m
		tab establishment_num allocation, m
		tab urbanrural_num allocation, m
		sum numberofpupils if allocation==0
		sum numberofpupils if allocation==1

		restore

		//Pupil level characteristics

		//year 5
		preserve
		keep if year_num==5
		di in red "Year 5"
		destring fsmeligible_aut18, replace
		tab fsmeligible_aut18 allocation, m
		tab female allocation, m

		egen z_ks1_mathsread = std(ks1_mathsread)

			sum ks1_mathsread if allocation==0
			sum z_ks1_mathsread if allocation==0

			sum ks1_mathsread if allocation==1
			sum z_ks1_mathsread if allocation==1

		egen z_agem = std(agem)

		sum agem if allocation==0
		sum z_agem if allocation==0

		sum agem if allocation==1
		sum z_agem if allocation==1

		restore

		//year 6
		preserve
		keep if year_num==6
		di in red "Year 6"
		destring fsmeligible_aut17, replace
		tab fsmeligible_aut17 allocation, m
		tab female allocation, m

		egen z_ks1_mathsread = std(ks1_mathsread)

			sum ks1_mathsread if allocation==0
			sum z_ks1_mathsread if allocation==0

			sum ks1_mathsread if allocation==1
			sum z_ks1_mathsread if allocation==1

		egen z_agem = std(agem)

		sum agem if allocation==0
		sum z_agem if allocation==0

		sum agem if allocation==1
		sum z_agem if allocation==1

		restore
*/

	//all of the main analysis from here down -- primary, secondary and mean imputation
(subgroups are separate)
	eststo clear
	//primary and secondary analysis
	global outcomes "ks2_mathsread ks2_matmrk ks2_readmrk"
	foreach y of global outcomes {
	*Year Split
		forvalues i = 5/6 {
			eststo `y'x1`i': regress `y' $x1 if ///
			year_num==`i', cluster (school_urn)
			summ `y' if e(sample)==1 & allocation==0
			estadd loc controlmean = string(`r(mean)', "%9.3f")
			estadd loc n = string(e(N), "%9.0fc")

			*To produce effect sizes and other information required for EEF tables
			sum `y' if allocation==0 & year_num==`i'
			 local n0 = r(N)
```

```
                        local mean0 = r(mean)
                        local sd0 = r(sd)
                        sum `y' if allocation==1 & year_num==`i'
                        local n1 = r(N)
                        local mean1 = r(mean)
                        local sd1 = r(sd)

                        local sstar = (((`n1'-1)*(`sd1'^2) + (`n0'-1)*(`sd0'^2))/(`n1' + `n0' -
2))^.5
                        local J = exp(lngamma((`n1' + `n0' - 2)/2) - ln(((`n1' + `n0' - 2)/2)^.5) -
lngamma((`n1' + `n0' - 2 - 1)/2))

                        local g   = `J' * _b[allocation] / `sstar'                      // Hedges g
                        local lbg = `J' * (_b[allocation] - 1.96*_se[allocation]) / `sstar'      //
Hedges g lower 95% confidence interval
                        local ubg = `J' * (_b[allocation] + 1.96*_se[allocation]) / `sstar'      //
Hedges g upper 95% confidence interval

                        di `n0'
                        di `n1'
                        di `sd0'
                        di `sd1'
                        di `g'
                        di `sstar'
                       di `lbg'
                       di `ubg'

                       mean `y' if allocation==1 & year_num==`i'
                       mean `y' if allocation==0 & year_num==`i'

               }



       //Sensitivity analysis using mean imputation for ks1 attainment -- primary and secnodary
specification (probs just report primary)
       *Year Split
               forvalues i = 5/6 {
                       eststo `y'impute`i': regress `y' $xi1 if ///
                       year_num==`i', cluster (school_urn)
                       summ `y' if e(sample)==1 & allocation==0
                       estadd loc controlmean = string(`r(mean)', "%9.3f")
                       estadd loc n = string(e(N), "%9.0fc")

               /*
                       eststo `y'x2`i': regress `y' $xi2 if ///
                       year_num==`i', cluster (school_urn)
                       summ `y' if e(sample)==1 & allocation==0
                       estadd loc controlmean = string(`r(mean)', "%9.3f")
                       estadd loc n = string(e(N), "%9.0fc")
               */
               }
       }
       //

//SUBGROUP ANALYSIS - looping over same thing again, for separate years -- loop over FSM status


               *Year Split
                       forvalues i = 5/6 {
                              eststo ks2_mathsreadx1`i'fsm: regress ks2_mathsread $x1 if ///
                              year_num==`i' & everfsmspryr==1, cluster (school_urn)
                              summ ks2_mathsread if e(sample)==1 & allocation==0
                              estadd loc controlmean = string(`r(mean)', "%9.3f")
                              estadd loc n = string(e(N), "%9.0fc")


                       *To produce effect sizes and other information required for EEF tables
                       sum ks2_mathsread if allocation==0 & year_num==`i'
                        local n0 = r(N)
                        local mean0 = r(mean)
                        local sd0 = r(sd)
                        sum ks2_mathsread if allocation==1 & year_num==`i'
                        local n1 = r(N)
                        local mean1 = r(mean)
                        local sd1 = r(sd)

                        local sstar = (((`n1'-1)*(`sd1'^2) + (`n0'-1)*(`sd0'^2))/(`n1' + `n0' -
2))^.5
```

```
                        local J = exp(lngamma((`n1' + `n0' - 2)/2) - ln(((`n1' + `n0' - 2)/2)^.5) -
lngamma((`n1' + `n0' - 2 - 1)/2))

                        local g   = `J' * _b[allocation] / `sstar'                      // Hedges g
                        local lbg = `J' * (_b[allocation] - 1.96*_se[allocation]) / `sstar'      //
Hedges g lower 95% confidence interval
                        local ubg = `J' * (_b[allocation] + 1.96*_se[allocation]) / `sstar'      //
Hedges g upper 95% confidence interval

                        di `n0'
                        di `n1'
                        di `sd0'
                        di `sd1'
                        di `g'
                        di `sstar'
                       di `lbg'
                       di `ubg'

                       mean ks2_mathsread if allocation==1 & year_num==`i'
                       mean ks2_mathsread if allocation==0 & year_num==`i'

                       }

       estout _all using "all.xls", replace cells(b(fmt(3)star) se(par(`"="("' `")""') fmt(3)))
starlevels(+ 0.10 * 0.05 ** 0.01) ///
       stats(N controlmean)
       *, labels("M. of obs." "Control Mean")

restore

*******************************************************************************
*******************Balance Checks at baseline for analysed groups**************
*******************************************************************************
/*
Balance check on:
       - baseline KS1 reading and maths point scores
       - proportion of students who are female
       - proportion of student ever eligible for FSM
       - proportion of students for whom English is an additional language
       - Age of student in months
Using absolute standardised differences
*/
//Calculating absolute standardised differences only for continuous variables
//(ks1 baseline and age) -- not doing this for categorical variables
//english as first language is not present in data
       global indbal "ks2_ks1matps ks2_ks1readps agem female everfsmspryr"
       forvalues i = 5/6 {
              foreach y of global indbal {
                     forvalues j = 0/1 {
                            summ `y' if year_num==`i' & allocation==`j'
                            gen `y'`i'_mean_t`j'=r(mean)
                            gen `y'`i'_sd_t`j'=r(sd)
                            gen z_`y'`i'_t`j' = abs(`y' - `y'`i'_mean_t`j') / `y'`i'_sd_t`j' if
year_num==`i' & allocation==`j'
                     }
                            ztest z_`y'`i'_t0 == z_`y'`i'_t1 if year_num==`i', unpaired
              }
       }

       //for the rest of the categorical variables // couldn't get EAL
       global categoricalbalance "female everfsmspryr"
       forvalues i = 5/6 {
              foreach y of global categoricalbalance {
                     regress `y' allocation if year_num==`i', cluster(school_urn)

              }
       }

       //

**************************************************
**********Non-compliance with the intervention*********
**************************************************
/*CACE analysis to estimate intervention effects on treated children+

Caution the results, given likely endogeneity of attendance and motivation --
 should we be doing this then?

SAP says that compliance means the teacher submitted at least 10 recordings over
 the academic year and received at least 2 reports from VC
```

```
        - Melbourne sent this data back in 2 phases: Phase 1 and Phase2
        Phase 1 consisted of 3 different rounds. Within each round, teachers were
        expected to record 5 lessons -- they would then receive 5 transcripts and
        1 feedback report.

        Phase 2 - wanted to recruit only teacher who'd participated in phase 1, but
        struggled with recruitment so this round opened up to other teachers that
        didn't participate in Phase 1.
        This consisted of 2 rounds. Within each round, teachers were to upload 3
        transcripts in each round, would then receive 1 feedback report

        Define compliance such that teachers submitted either
        10 records across both Phases, or 5/6 of Phase 2
*/

        replace nameofschool=trim(nameofschool)
        replace teachername=trim(teachername)
        sort nameofschool
        split teachername,p(" " "/" "Mr" "Mrs" "Miss" "Ms" "+" ".")
        forvalues i = 1/9 {
                replace teachername`i'=trim(teachername`i')
        }
        gen iteachername=teachername1+" "+teachername2+" "+teachername3+" "+teachername4+"
"+teachername5+" "+teachername6+" "+teachername7+" "+teachername8+" "+teachername9
        replace iteachername=trim(iteachername)
        drop teachername1 teachername2 teachername3 teachername4 teachername5 teachername6
teachername7 teachername8 teachername9
        split iteachername,p("," " and " "&" "  ")
        replace iteachername1=trim(iteachername1)
        replace iteachername2=trim(iteachername2)
        replace iteachername3=trim(iteachername3)
        gen teachername1=iteachername1
        gen teachername2=iteachername2
        replace teachername2=iteachername3 if teachername2==""
        drop iteachername*
        replace teachername1=lower(teachername1)
        replace teachername2=lower(teachername2)
        replace nameofschool=lower(nameofschool)
        gen teachnm=teachername1 if teachername2==""
        replace teachnm=teachername1+" & "+teachername2 if teachername2!=""

merge m:1 randomid using "P:\Input\_Researcher file ingest\20190903 JW Ingest
1011227\survey_compliance_forONS.dta"



        global x1_comp  "ks1_mathsread dblock2-dblock9" //as intended and specified in SAP with
strata dummies

//import/merge teacher dataset and merge using student random ID
*xi is always imputed
*i 5/6 - year group
*j - definition of compliance: 1- 10 overall or 2-5 in phase 2,

eststo clear
forvalues i = 5/6 {
        *Year 5 & 6

                eststo compliancex1`i': ivregress 2sls ks2_mathsread $x1_comp (complier1 =
allocation)  if year_num==`i', cluster (school_urn) first
                        summ `y' if e(sample)==1 & allocation==0
                        estadd loc controlmean = string(`r(mean)', "%9.3f")
                        estadd loc n = string(e(N), "%9.0fc")


}

        estout _all using "noncompliance.xlsx", replace cells(b(fmt(3)star) se(par(`"="("' `")""')
fmt(3))) starlevels(+ 0.10 * 0.05 ** 0.01) ///
        stats(N controlmean)
        **need to also copy manually output of these regs because don't know how to export both
stages

        *estout _all using "noncompliance.xls", replace cells(b(fmt(3)star) se(par(`"="("' `")""')
fmt(3))) keep(*allocation*) starlevels(+ 0.10 * 0.05 ** 0.01) ///
        *stats(N controlmean)
/**/

**************************************************
*******************Additional Analysis***********
**************************************************
```

```
/*
Descriptive dose response analysis -- treatment variable is pseudo-continuous based on VC rubric
performance of treated teachers
- Need rubrics and scores from University of Melbourne merged onto teacher names
- regression: student outcomes on teacher performance

Conduct analysis using the Survey of Teacher Practice -- understand how self-reported changes in
teaching practice moderate any effect on attainment
*/

/*
Dose response analysis -- will only run for those with a number of report submissions
*/
eststo clear
        global outcomes "ks2_mathsread"
        foreach y of global outcomes {
        *Year Split
                forvalues i = 5/6 {
                        eststo `y'x1`i': regress `y' TSC ks1_mathsread dblock2-dblock9 if ///
                        year_num==`i', cluster (school_urn)

                }
        }

        estout _all using "doseresponse.xls", replace cells(b(fmt(3)star) se(par(`"="("' `")""')
fmt(3))) starlevels(+ 0.10 * 0.05 ** 0.01) ///
        stats(N)

/*
Impact on survey of teacher practice -- loop over each question as discussed earlier
*/
eststo clear

//cleaning up the survey questions a bit to do a more sensible analysis on this
egen q5_pre = rowmean(fQ5_1pre fQ5_2pre fQ5_3pre fQ5_4pre fQ5_5pre fQ5_6pre fQ5_7pre fQ5_8pre
fQ5_9pre fQ5_10pre fQ5_11pre fQ5_12pre fQ5_13pre fQ5_14pre fQ5_15pre fQ5_16pre)
egen q5_post = rowmean(fQ5_1post fQ5_2post fQ5_3post fQ5_4post fQ5_5post fQ5_6post fQ5_7post
fQ5_8post fQ5_9post fQ5_10post fQ5_11post fQ5_12post fQ5_13post fQ5_14post fQ5_15post fQ5_16post)
egen q10_pre = rowmean(fQ10_1pre fQ10_2pre fQ10_3pre fQ10_4pre fQ10_5pre fQ10_6pre fQ10_7pre
fQ10_8pre fQ10_9pre fQ10_10pre)
egen q10_post = rowmean(fQ10_1post fQ10_2post fQ10_3post fQ10_4post fQ10_5post fQ10_6post fQ10_7post
fQ10_8post fQ10_9post fQ10_10post)
egen q11_pre = rowmean(fQ11_1pre fQ11_2pre fQ11_3pre fQ11_4pre fQ11_5pre fQ11_6pre fQ11_7pre)
egen q11_post = rowmean(fQ11_1post fQ11_2post fQ11_3post fQ11_4post fQ11_5post fQ11_6post
fQ11_7post)

global post_avg "q5_post q10_post q11_post"
global pre_avg "q5_pre q10_pre q11_pre"



                forvalues i = 5/6 {
                eststo q5_y`i': regress ks2_mathsread q5_post q5_pre dblock2-dblock9 ks1_mathsread if
year_num==`i', cluster(school_urn)

                eststo q10_y`i': regress ks2_mathsread q10_post q10_pre dblock2-dblock9 ks1_mathsread
if year_num==`i', cluster(school_urn)

                eststo q11_y`i': regress ks2_mathsread q11_post q11_pre dblock2-dblock9 ks1_mathsread
if year_num==`i', cluster(school_urn)
                }

        estout _all using "surveyteacherpractice.xls", replace cells(b(fmt(3)star) se(par(`"="("'
`")""') fmt(3))) starlevels(+ 0.10 * 0.05 ** 0.01) ///
        stats(N)
```

Appendix K: Visible Classroom Teacher Practice Survey

Start of Block: Default Question Block

Q1
Thank you for participating in this survey about the Visible Classroom! We appreciate your time and input.

On the next page you will see a consent form that explains the study in more detail; we are required by the University of Melbourne to include this in all our surveys. Please read the consent form, indicate your consent if you choose to do so, then continue on to our survey which will begin on the following page.

Thank you!

Page Break

Q2
Research Study on the Impact of The Visible Classroom Project

Consent Form - Teacher Survey
Name of investigator(s): Dr Janet Clinton, Sarah Mason

Please read this consent form.

If you agree to the terms of consent, please select 'Next' at the bottom of the page to commence the survey. If you do not agree to the terms of consent, you may close this survey browser.

1. **I consent to participate in the project named above, the particulars of which – including details of the survey – have been explained to me.**
2. **I authorise the researchers to use for this purpose the survey referred to above.**
3. **I acknowledge that:**
    - **The possible effects of survey participation have been explained to me to my satisfaction.**
    - **I have been informed that I am free to withdraw from the project at any time without explanation or prejudice and to withdraw any unprocessed data previously supplied .**
    - **The project is for the purpose of research.**
    - **I have been informed that the confidentiality of the information**
    - **I provide will be safeguarded subject to any legal requirements.**
    - **I will not be named in any reports or presentations arising from the research.**

Page Break

Q3 Before we get started, please tell us a little about yourself.

Q4 1. What is your first name? (We only ask for your name so we can compare your answers in this survey to your answers at the end of the school year. No teacher will ever be named in any of our research).

_____

Q5 2. What is your last name?

_____

Q6 3. Which school do you work at?

_____

Q7 4. What year level do you currently teach?

- **o Year 5 (1)**
- **o Year 6 (2)**
- **o Other (Please Specify) (3) _____**

Page Break

Q8 First, we'd like to learn a little about your current teaching practice.

Q9 5. How frequently do you currently use the following teaching strategies?

	Never (1)	Sometimes (2)	About half the time (3)	Most of the time (4)	Always (5)
Promote deep understanding (1)	o	o	o	o	o
Promote collaboration (2)	o	o	o	o	o

Help students make connections to the real world (3)	o	o	o	o	o
Check fact recall by asking convergent (closed-ended) questions (4)	o	o	o	o	o
Review previously learned concepts (5)	o	o	o	o	o
Repeat comments and questions to encourage and reinforce correct answers (6)	o	o	o	o	o
Create a positive classroom environment (7)	o	o	o	o	o
Provide step-by-step instructions on completing tasks/activities (8)	o	o	o	o	o
Use prompting or probing (9)	o	o	o	o	o
Conclude lessons by summarising key points (10)	o	o	o	o	o
Set clear behaviour expectations (11)	o	o	o	o	o
Introduce and explain new vocabulary and concepts (12)	o	o	o	o	o
Use written/visual/audio resources to	o	o	o	o	o

support learning (13)					
Emphasise important points (14)	o	o	o	o	o
Provide immediate, specific and corrective feedback to students (15)	o	o	o	o	o
Clearly express teaching goals and refer to those goals throughout the lesson (16)	o	o	o	o	o

Page Break

Q10 To help us understand more about you and your context, we would like to learn about your thoughts on feedback in general, and your school.

Q11 6. How much do you disagree or agree with the following statements:

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
I feel self-assured when dealing with feedback (1)	o	o	o	o	o

Compared to others, I am more competent at handling feedback (2)	o	o	o	o	o
I believe that I have the ability to deal with feedback effectively (3)	o	o	o	o	o
I feel confident when responding to both positive and negative feedback (4)	o	o	o	o	o
I know that I can handle the feedback that I receive (5)	o	o	o	o	o
Feedback contributes to my success at work (6)	o	o	o	o	o
To develop my skills at work I rely on feedback (7)	o	o	o	o	o
Feedback is critical for improving performance (8)	o	o	o	o	o
Feedback from other educators can help me advance in a school (9)	o	o	o	o	o

I find that feedback is critical for reaching my goals (10)	o	o	o	o	o

Q12 7. Thinking about your school, how much do you disagree or agree with the following statements:

	Strongly disagree (1)	Somewhat disagree (2)	Neither agree nor disagree (3)	Somewhat agree (4)	Strongly agree (5)
In my school, people are rewarded for learning (1)	o	o	o	o	o
In my school, people spend time building trust with each other (2)	o	o	o	o	o
In my school, teams/groups revise their thinking as a result of group discussions or information collected (3)	o	o	o	o	o
My school makes its lessons learned available to all employees (4)	o	o	o	o	o
My school recognizes people for taking initiative (5)	o	o	o	o	o

My school works together with the outside community to meet mutual needs (6)	o	o	o	o	o
In my school, leaders continually look for opportunities to learn (7)	o	o	o	o	o

Page Break

Q13 We're also interested in your expectations for the Visible Classroom project overall.

Q14 8. Based on what you know so far, how much do you disagree or agree with the following statements:

Being involved in Visible Classroom would….

	Strongly Disagree (9)	Disagree (10)	Neutral (11)	Agree (15)	Strongly agree (16)
Positively impact teaching practice (1)	o	o	o	o	o
Allow identification new goals for teacher's professional development (2)	o	o	o	o	o

Page Break

Q15 Thank you for your help so far! So that we know more about you, please answer these brief demographic questions.

Q16 9. At which type of schools have you taught during your career? Select all that apply.

- ☐ **Academy (1)**
- ☐ **Free (2)**
- ☐ **Private (3)**
- ☐ **Other (Please Specify) (4) _____**

Q17 10. What was the name of your initial teaching qualification?

- **o Bachelor of Education (1)**
- **o Post Graduate Certificate in Education (2)**
- **o School-Centred Initial Teacher Training (3)**
- **o School Direct (4)**
- **o Teach First (5)**
- **o Other (please specify) (6) _____**

Q18
11. Approximately how many years of teaching experience do you have?

- **o 1-3 (1)**
- **o 4-6 (2)**
- **o 7+ (3)**

Page Break

Q19 Thank you very much for your participation!

End of Block: Default Question Block

Teacher Practice Survey: Overview of constructs

Overarching research questions

1. **To what extent do teachers adapt their teaching practice in response to evaluative feedback?**
2. **What individual and organisational factors affect whether teachers will adapt their behaviour in response to evaluative feedback?**
    a. **Hypothesis 1: Feedback self-efficacy will be positively related to behaviour change [alt: feedback self-efficacy will moderate the relationship between evaluative feedback and behaviour change]**
    b. **Hypothesis 2: Feedback utility will be positively related to behaviour change [alt: feedback utility will moderate the relationship between evaluative feedback and behaviour change]**
    c. **Hypothesis 3: Quality of the mentor relationship will be positively related to behaviour change [alt: quality of mentor relationship will moderate the relationship between evaluative feedback and behaviour change]**
    d. **Hypothesis 4: Organizational learning climate will be positively related to behaviour change [alt: Organizational learning climate will moderate the relationship between evaluative feedback and behaviour change]**
3. **To what extent does feedback orientation (i.e. feedback self-efficacy + feedback utility) change over time?**
4. **What are the relationships between individual, organisational and feedback reaction constructs?**

Construct			Items
Feedback self-efficacy **Linderbaum, B. A., & Levy, P. E. (2010). The development and validation of the Feedback Orientation Scale (FOS). Journal of Management, 36(6), 1372-1405.**	**An individual's tendency to have confidence in dealing with feedback situations and feedback.**	**α=.80, 78 (study 1, 2)** **Test-retest = .6** **Convergent validity = general self-efficacy (r = .46); locus of control (r = -.25)** **Discriminant validity = public self-consciousness (r = -.16)** ****See unique contributions in table pasted below**	1. **I feel self-assured when dealing with feedback** 2. **Compared to others, I am more competent at handling feedback** 3. **I believe that I have the ability to deal with feedback effectively.** 4. **I feel confident when responding to both positive and negative feedback** 5. **I know that I can handle the feedback that I receive**

Feedback utility **Linderbaum, B. A., & Levy, P. E. (2010). The development and validation of the Feedback Orientation Scale (FOS). Journal of Management, 36(6), 1372-1405.**	**An individual's tendency to believe that feedback is instrumental in achieving goals or obtaining desired outcomes at work.**	**α=.86,.88 (study 1, 2)** **Test-retest = .6** **Convergent Validity = job involvement (r = .17), feedback quality (r = .41)** **Discriminant validity = fear of negative evaluation (r = .00)** ****See unique contributions in table pasted below**	1. **Feedback contributes to my success at work** 2. **To develop my skills at work I rely on feedback.** 3. **Feedback is critical for improving performance.** 4. **Feedback from supervisors can help me advance in a company [change to: Feedback from other educators can help me advance in a school]** 5. **I find that feedback is critical for reaching my goals.**

Org Learning ****7-item version of the Dimensions of Organizational Learning Scale** **Yang, B. (2003). Identifying valid and reliable measures for dimensions of a learning culture. Advances in Developing Human Resources, 5(2), 152-162.**	**Measure's the abstract construct of a learning culture.**	**α=.84**	**1. In my school, people are rewarded for learning** **2. In my organization, people spend time building trust with each other** **3. In my organization, teams/groups revise their thinking as a result of group discussions or information collected.** **4. My organization makes its lessons learned available to all employees** **5. My organization recognizes people for taking initiative.** **6. My organization works together with the outside community to meet mutual needs** **7. In my organization, leaders continually look for opportunities to learn.**

Appendix L: Letter of Ethical Approval

THE BEHAVIOURAL INSIGHTS TEAM

Behavioural Insights Team External Ethical Advisory Panel

Human Research Ethics Committee

4th June 2016

Dear Aisling,

Trial Number: 2015146

Trial Name: EEF Visible Classroom Evaluation

Thank you for your submission to the panel. Following our review, comments and confirmation of your revised documentation, I am pleased to confirm a favourable ethical opinion of the above research on the basis described in your trial protocol and accompanying documentation. This letter provides ethical clearance for the launch of the trial within the next three years, i.e. until 04/06/2019. If the trial is to be launched after this date, we request that the panel be informed as reapplication may be necessary.

If prior to launch, a substantial amendment is required, the panel should be notified for further review.

With the committee's best wishes for the success of the project,

Prof Peter John

External Ethical Advisory Panel chairperson

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

https://educationendowmentfoundation.org.uk
@EducEndowFoundn
Facebook.com/EducEndowFoundn