Education Endowment Foundation

First Thing Music

Evaluation Report

September 2021

Kimberly Bohling, Matthew Barnard, Lauren Crouch, Alexander Whitefield, Bridie Murphy, Fabian Gunzinger, Jake Anders, Nikki Shure, Dominic Wyse

UCL Institute of Education

THE BEHAVIOURAL INSIGHTS TEAM

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 **4**
Executive summary **5**
Introduction **7**
Methods **17**
Impact evaluation **42**
Implementation and process evaluation (IPE) **53**
Conclusion **72**
References **77**
Appendix A: EEF cost rating **79**
Appendix B: Security classification of trial findings **80**
Appendix C: Effect size estimation **82**
Further appendices **83**

About the evaluator

The project was independently evaluated by a team from University College London (Dr Jake Anders, Dr Nikki Shure, Professor Dominic Wyse, Professor John Jerrim, Professor Gemma Moss and Professor Andrew Burn) and the Behavioural Insights Team (Kimberly Bohling, Dr Matt Barnard, Lauren Crouch, Alexander Whitefield, Bridie Murphy and Fabian Gunzinger).

The lead evaluators were Kimberly Bohling and Dr Matt Barnard.

BIT contact details:

Kimberly Bohling

Behavioural Insights Team

4 Matthew Parker Street

London SW1H 9NP

Email: **info@bi.team**

Executive summary

The project

First Thing Music (FTM) is a programme inspired by the Kodály method of music instruction, which provides a structured, sequential music curriculum of increasing progression. The intervention was designed to be delivered by teachers, with support from specialist music practitioners (external to the school), to classes of Year 1 pupils (aged 5–6) for 15 minutes at the start of each school day for one academic year. The programme was developed by the Tees Valley Music Service.

This evaluation tested the efficacy of FTM through a two-arm clustered randomised controlled trial, including 3004 pupils from 64 schools. The trial measured the impact of the project on pupils' reading attainment, creative self-efficacy (confidence in one's own creative abilities) and social skills. In addition, survey data was collected from all schools, and case studies involving interviews and classroom observations were collected from six schools. This evaluation was jointly funded by the Education Endowment Foundation (EEF) and the Royal Society of Arts (RSA). The trial started in June 2018 and ended in July 2019.

Key conclusions
1. Pupils in the First Thing Music (FTM) intervention group made the equivalent of one month's additional progress in reading, on average, compared to pupils in the control group. This is our best estimate of impact which has a low to moderate security rating. However, as with any study, there is uncertainty around the result: the possible impact of this programme ranges from no additional progress to positive effects of three additional months of progress.
2. Pupils in the FTM intervention group achieved lower scores in an assessment of social skills than pupils in the control group at the end of the trial, though the difference in scores was small and there is some uncertainty around this result. There was no evidence from the impact evaluation that the FTM intervention had an impact on creative self-efficacy.
3. Only 40 percent of teachers in the intervention group delivered at least 80 percent of the possible music sessions and attended at least four of the six training sessions. The impact of the programme on reading attainment was higher, on average, in classrooms where this threshold for training and delivery was met.
4. Around half of the teachers (54 percent) delivered the programme 'first thing' as planned, while other teachers moved the sessions to later in the day. Timetabling difficulties, such as the sessions overrunning, were a barrier to delivery in the morning.
5. Teachers perceived the programme to have had a positive impact on pupils' musical skills, social skills, creativity and self-regulation, but more than half of those surveyed felt that the programme was unlikely to impact reading attainment.

EEF security rating

These findings have a low to moderate security rating. This was an efficacy trial, which tested whether the intervention worked under developer-led conditions in a number of schools. The trial was a well-designed two-arm clustered randomised controlled trial and was well powered.

28 percent of pupils who started the trial were not included in the final analysis due to six schools not participating in testing, and due to pupils changing school, being absent on testing dates or refusing testing. Pupils eligible for free school meals (FSM) were more likely to be missing from the final analysis than other pupils. Pupils in the control group were also more likely to be missing from the final analysis than intervention group pupils. Furthermore, the intervention was not delivered as intended in many schools, with only 40 percent of teachers meeting the training and delivery compliance threshold and only 54 percent delivering the music sessions 'first thing' in the morning. These issues make it harder to provide a secure estimate of the size of the impact on the pupils in the trial.

Additional findings

The effect of the FTM programme on pupils eligible for FSM was lower, on average, than for other pupils and equivalent to 0 months' additional progress in reading. However, this finding is based on a small sub-sample of pupils and the observed difference in reading outcomes was small, so we cannot be certain of whether the programme had a different impact for pupils eligible for FSM.

73 percent of teachers attended at least four of the six FTM training sessions, but only 42 percent delivered at least 80 percent of the possible music sessions to their class. Overall, this meant that 40 percent of teachers completed the minimum expected amount of training and delivery, which was assessed using administrative data collected by the developer. Notably, in classrooms where this threshold for training and delivery was met, the average effect of the programme on reading attainment was twice as large as for the intervention group as a whole. This finding should be interpreted with caution, as it is based on a small, non-random sub-sample of the most engaged classrooms, but it could indicate that higher engagement with FTM could lead to greater gains in reading attainment.

Teachers perceived the programme to have a positive impact on pupils' musical skills, social skills, creativity and self-regulation, but more than half of those surveyed felt the programme was unlikely to impact reading attainment. This contrasts with the impact evaluation results, which found no evidence that the programme had a positive effect on social skills or creative self-efficacy but did find some evidence to suggest that pupils in the intervention group made an average of one month's additional progress in reading compared to the control group. Some teachers also felt that FTM would be more likely to impact maths than reading attainment because of the numerical patterns and counting involved in the sessions.

Wider research on the impact of music education on attainment shows some promise, but few studies in this area have used rigorous experimental methods. The findings of this evaluation provide tentative evidence to support the hypothesis that music education can contribute to improvements in children's reading attainment. However, further research is needed to strengthen the evidence base and improve understanding of the relationship between music education and reading skills development.

Cost

The average cost of FTM for one school is around £1845, or £52 per pupil per year when averaged over three years.

Impact

Table 1: Summary of impact on primary outcome

Outcome / group	Effect size[1] (95% confidence interval)	Estimated months' progress	EEF security rating	Total no. of pupils (treatment, control)	*p*-value	EEF cost rating
Reading (Year 1)	0.07 (−0.02,0.16)	1	🔒🔒🔓🔓🔓	2150 (1116, 1034)	0.13	**£** £ £ £ £
Reading (Year 1, free school meals)	0.00 (−0.20, 0.19)	0	N/A	451 (245, 206)	0.97	**£** £ £ £ £

[1] Intent to treat (ITT) Hedges' *g*

Introduction

Background

This evaluation is part of a round of funding between the Education Endowment Foundation (EEF) and the Royal Society of Arts (RSA), to test the impact of different cultural learning strategies in English schools entitled 'Learning about Culture'.[2] The aim is to improve the evidence base around arts-based education programmes. It consists of five programmes: two in Key Stage 1 (KS1) (Years 1 and 2) and three in Key Stage 2 (KS2) (Year 5). Despite the unique aspects of these intervention models, there are many similarities in how they are delivered and what they hope to achieve.[3] The programmes have been supported by Arts Council England.

The background for the study is that focusing on increasing attainment in literacy and numeracy has been criticised for marginalising art, music and cultural studies in English schools (Warwick Commission, 2015). The UK Government's Culture and Sport Evidence review (Newman et al., 2010), which summarised much of the observational and qualitative research in this area, showed pupil participation in cultural learning programmes (from piano training to theatre-based drama projects) to be correlated with higher levels of achievement in mathematics and literacy / English in both primary and secondary school. The review also linked participation in cultural learning programmes to faster language development in the early years and improved cognitive ability. Additionally, large cohort observational studies in the US have suggested that correlations between mathematics and literacy gains and participation in arts classes and extracurricular activities are particularly large for pupils from low-income groups (Catterall, 2009; Catterall et al., 2012). This evidence suggests that cultural learning correlates with academic attainment; however, the causal nature of this relationship remains unclear – a key motivating factor for the 'Learning about Culture' programme.

The First Thing Music (FTM) intervention aims to improve children's music understanding and their reading and social skills through daily 15-minute music sessions at the start of the school day. There are a number of theories as to how music instruction might improve literacy skills and these causal pathways are referred to as 'transfer' (Cogo-Moreira et al., 2013). Transfer refers to the ability to use a skill or knowledge developed in one context in another context (Barnett & Ceci, 2002). The link between music education and improved reading skills can be considered 'far transfer' because the two are not directly related. The development of phonological awareness is cited as one potential aspect linking music and reading development (Patscheke et al., 2016). Phonological awareness is the awareness of and ability to manipulate sound in spoken communication, and it is considered a foundational skill to literacy. Activities like singing may facilitate increased phonological awareness through practice with rhyming, alliteration and syllable segmentation. Non-experimental studies have found positive correlations between music education and phonological awareness (Anvari et al., 2002; Gromko, 2005; Foregard et al., 2008; Patscheke et al., 2016). A meta-analysis of 13 studies exploring the impact of music education on phonological awareness in children found an effect size of $d = 0.20$ (Gordon et al., 2015). However, the strength of this evidence is weak. Although all studies utilised a control group, only six used random assignment; and more than half had a sample size smaller than 50 pupils.

FTM would be considered to have a far transfer relationship to improved literacy skills, as the programme seeks to improve music understanding; music education is hypothesised to improve phonological awareness, and phonological awareness is a core aspect of literacy development. When considering the far transfer between music education and improved pupil attainment and social–emotional skills, the evidence is mixed and weak. In a review of arts education research, 30 studies were identified as exploring music education in primary schools (See & Kokatsaki, 2016). Of these studies, several reported positive effects on reading and maths skills. However, none were deemed to have produced strong evidence, five produced medium–weak evidence and the remainder were deemed weak. Studies were considered weak due to a variety of factors, which were largely non-experimental designs and small sample sizes.

[2] https://www.thersa.org/globalassets/pdfs/reports/rsa-learning-about-culture-report.pdf

[3] For an overarching flow diagram of the programme similarities, please see Appendix D.

FTM is inspired by the Kodály method of music instruction, which provides a structured, sequential music curriculum of incremental progression. Of the studies in See & Kokatski's (2016) review, those using the Kodaly approach showed weak evidence of positive impact on particular mathematics skills (*n* = 67 pupils in Courey et al., 2012) and creativity (*n* = 48 pupils in Duncan, 2007), and moderate evidence of impact on speech segmentation skills (*n* = 24 pupils in François, 2013). A small pilot study (*n* = 56 pupils) of the Kodály method, which informed the development of FTM, found positive improvements in behaviour, writing and numeracy (See & Ibbotson, 2018).

The available evidence on the impact of music education shows promise, but there has been little use of rigorous experimental methods which can provide the strongest evidence of impact. Therefore, the current evaluation of FTM comprises an impact evaluation and implementation and process evaluation (IPE) to provide an in-depth understanding of the programme's impact, including how and why the hypothesised outcomes have or have not been achieved. The primary aim of the impact evaluation is to test whether FTM impacts pupils' reading attainment and social skills over the course of one school year for Year 1 pupils. The IPE aims to identify the mechanisms that are present in bringing about the change identified from the impact evaluation, and any factors that have influenced such impact and delivery of the programme as intended.

A note on protocol deviations related to data sharing

We note that it has been necessary for the analysis of this trial to deviate substantially from the initial plans set out in the project protocol and statistical analysis plan (SAP).[4] This stems from issues in accessing the baseline data that were to have been obtained from the Department for Education's (DfE) National Pupil Database (NPD). During the implementation of the trial, the DfE changed the way in which data from the NPD are made available to researchers, switching from providing extracts that can be used alongside project data within evaluator's own secure computing systems to requiring access within the Office for National Statistics (ONS) Secure Research Service (SRS). In turn, this means that it is now necessary for project data to be uploaded to the SRS. Given that this project data is considered personal data over which the evaluators are data controllers, this requires the conclusion of an appropriate data sharing or processing agreement between the evaluator and the DfE and/or the ONS in order to provide legally required reassurance by the DfE/ONS about the treatment of personal data over which the evaluator is controller. This implication of altering their processes appears not to have been fully planned for by the DfE and ONS and, as such, attempting to conclude such an agreement that does not appear to have precedents has been subject to extended negotiations and delays between the evaluators and DfE/ONS, which have been severely exacerbated by additional workload for these organisations due to the COVID-19 pandemic. In the interests of completing these evaluations, and after discussion with the EEF and project teams, the decision was made to proceed with analysis with deviations from protocol flagged throughout the *Methods* section. The main change is the use of pupils' FSM and EAL status as baseline measures, instead of Early Years Foundation Stage Profile (EYFSP) scores. These deviations were agreed with the EEF ahead of conducting the analysis. Beyond issues inherent in deviating from pre-registered protocol, the main implication for the analysis is reduction in statistical power relative to expectations. This means that the impact evaluation is not able to statistically detect effects as small as expected in the SAP.

It is important to understand the implications of this change. The purpose of including baseline measures in the current evaluation is to increase its statistical precision (i.e., to reduce the uncertainty around intervention impact estimates, which makes them more likely to be statistically significant). Importantly, both the original and the substituted baseline measures are taken from prior to the randomisation and intervention. Therefore, due to the randomised nature of the evaluation, their inclusion does not bias any intervention impact estimates, but only affects the statistical uncertainty around these estimates (i.e., the extent to which they are detectable as statistically significant). Given EEF policy to report impact estimates whether or not they are statistically significant, there is an increased risk that headline positive or negative effects are just due to this uncertainty, rather than representing a true effect. As a result, we particularly stress the importance of statistical significance as a check on interpretation of the results in this report.

[4] First Thing Music Evaluation Protocol; First Thing Music Statistical Analysis Plan (SAP)

Intervention

The FTM programme is designed to improve children's musical understanding, reading and social skills by providing them with daily music lessons. The programme is closely aligned with the British Kodály Academy and its methodology, and is a relatively new programme (developed in 2017), which builds on the more established Kodály method (See & Ibbotson, 2018). The description below follows the Template for Intervention Description and Replication (TIDiER) checklist.[5]

1. **Brief Name**. First Thing Music (FTM)
2. **Why? (Rationale, theory, or goal).** The First Thing Music (FTM) programme aims to improve children's music understanding, and reading and social skills, through daily, 15-minute music sessions over the course of three terms. During the sessions, Year 1 pupils (aged 5–6) are introduced to basic musical principles through singing and movement activities, focusing mainly on steady beat, rhythm and pitch. The programme is inspired by the Kodály method of music instruction, which provides a structured, incrementally progressive, sequential music curriculum. Pupils are introduced to musical concepts, first through singing, clapping and games, before progressing to musical symbols and rhythmic notation. Learning to detect rhythm and count beats and engaging with activities related to phonological awareness, such as rhyming, are theorised to support the development of numeracy and literacy skills. The intervention also hopes to develop children's social skills through its focus on collaborative games and activities, which promote turn-taking and teamwork.
3. **What (materials).** Teachers have access to an online video resource, via the programme website, with recordings of all 'out of setting' continuing professional development (CPD) sessions. Teachers also have access to a resource booklet, which contains a written introduction to the Kodály-based approach, as well as songs, rhymes and games, in both visual and recorded format.
4. **What (procedures).** In daily, 15-minute music sessions, pupils learn and practise songs and chants which gradually introduce the concepts of beat, rhythm and pitch. The programme takes a three-step approach, referred to as preparation, presentation and practice. In the preparation phase, children are guided by the teacher to develop an awareness of rhythm, beat and pitch, following specific activities that progress incrementally. For example, while introducing beat, children learn songs in which they move together with others to experience a steady beat in hands and fingers, then upper body, whole body, walking and finally on instruments. Here, children are learning as a group to identify and follow a beat, without knowing about beat as a concept.
   Teachers visually assess whether enough children can perform the activity before moving on to the presentation phase, in which children are introduced to each concept through verbal explanation and activities that enforce it visually. Continuing using beat as an example, this can include matching rhymes to visual patterns, or following and tapping out visual representations, such as 'heartbeats' on paper whilst listening to a song. These actions seek to allow children to feel, see and hear beat concurrently, and reinforces the sense of beat.
   Finally, in the practice phase, the songs on which the programme is based are repeated, with the level of challenge gradually increased, allowing children to develop a deeper understanding of beat, rhythm and pitch.
5. **Who (recipients).** FTM is developed to be used with Early Years and KS1 pupils of all abilities. It was only delivered to Year 1 pupils in this evaluation.
6. **Who (implementers).** The intervention is implemented by classroom teachers, with the support of external music practitioners (specialists), recruited by the British Kodály Academy Training Team, Tees Valley Music Service and the programme developers. The specialists provide training and mentoring to teachers throughout the duration of the programme. The specialists are required to have music training to degree or diploma level, a background in the Kodály methodology, and experience with whole-class teaching. Upon hiring, they received a day of training from the FTM team about the intervention and their role in its implementation.
   The training schedule is designed to gradually transfer implementation leadership from the specialist to the teacher, with the specialist resuming the lead role when new elements are introduced. The training schedule is as follows: before beginning the programme, teachers attend a day-long training session. In the first week of the

[5] https://www.bmj.com/content/348/bmj.g1687

intervention, the daily music sessions are led by the specialists. In week two, teachers lead the sessions themselves, without the specialist present. In week three, the specialist returns to team-teach alongside the teacher, supporting the teacher in taking the lead on some activities, and developing the sessions further. From week four onwards, the specialist visits once per week. In addition, the British Kodály Academy, supported by the specialists, provides two afternoon CPD sessions per term.

7 **How (mode of delivery).** Sessions are delivered face-to-face by teachers daily, occasionally with the support of specialists (as described in **6** above) to whole-class groups of Year 1 pupils. Ideally, the intervention is delivered at the start of the school day.

8 **Where (setting).** Sessions are delivered as 'carpet' sessions in the classroom or other available space. The space should enable face-to-face group work, as opposed to all pupils facing the teacher.

9 **When and how much (dosage).** FTM sessions last 15 minutes and are delivered daily. This continues over the course of three terms (one academic year). If delivered daily as intended, pupils receive approximately 47 hours of the programme.

The dosage was informed by the developer's pilot study (See & Ibbotson, 2017). The developer's rationale for 'a little and often' was that it is not only effective for the children, but also easier for teachers to implement than full music lessons of 40–60 minutes. The developer found in the pilot that the 'first thing' element was effective in supporting a good start to the school day and helping the children to focus thereafter.

Sessions are ideally delivered at the start of the day, where possible, which may require teachers to adjust their timetables to accommodate the sessions.

10 **Tailoring.** The intervention was not planned to be personalised, titrated or adapted. Early in the development of the intervention, the team developed ways of supporting pupils with additional needs, such as those with English as an additional language (EAL) status, or challenging behaviour. Teachers receive guidance on this at the training sessions and further support is available on the FTM website.[6] At the point of evaluation, the programme was not fully adapted to work with children with visual or hearing impairments; however, this is in development.

11 **Modifications.** The developer did not make any modifications to the interventions. However, as teachers were primarily responsible for delivery, teachers occasionally made modifications, which are detailed as findings in the *IPE results*.

12 **How well (planned).** There are two implementation dimensions – compliance and implementation fidelity.

The definition of compliance was agreed with the FTM team. Per the evaluation protocol, sufficient compliance to the intervention was considered to be delivering 80 percent of all possible sessions and the implementing teacher missing no more than two of the offered training sessions. The training programme consisted of one full-day training in September and five afternoon sessions (October, November, January, February and May).[7] Teachers needed to attend at least four of the six training sessions offered, to meet the minimum compliance threshold. sessions. Where staff were struggling to deliver the intervention, FTM staff provided support. Teachers were provided with half-termly logs to record their delivery of the programme. FTM collected the completed logs at the end of each half term.

Implementation fidelity was conceptualised as how the way in which the intervention was implemented in practice compares to the intended implementation of the intervention as described in this section.

Logic model

The intervention's core activity consists of daily, 15-minute music sessions over the course of three terms (see Figure 1 for the original logic model). Sessions are enabled and supported by the Kodály methodology booklet; an initial training session with a Kodály specialist; support from a primary music specialist in weeks one and three; five after-school sessions; and participation from the head teacher in at least one session per term. The FTM programme aims to improve

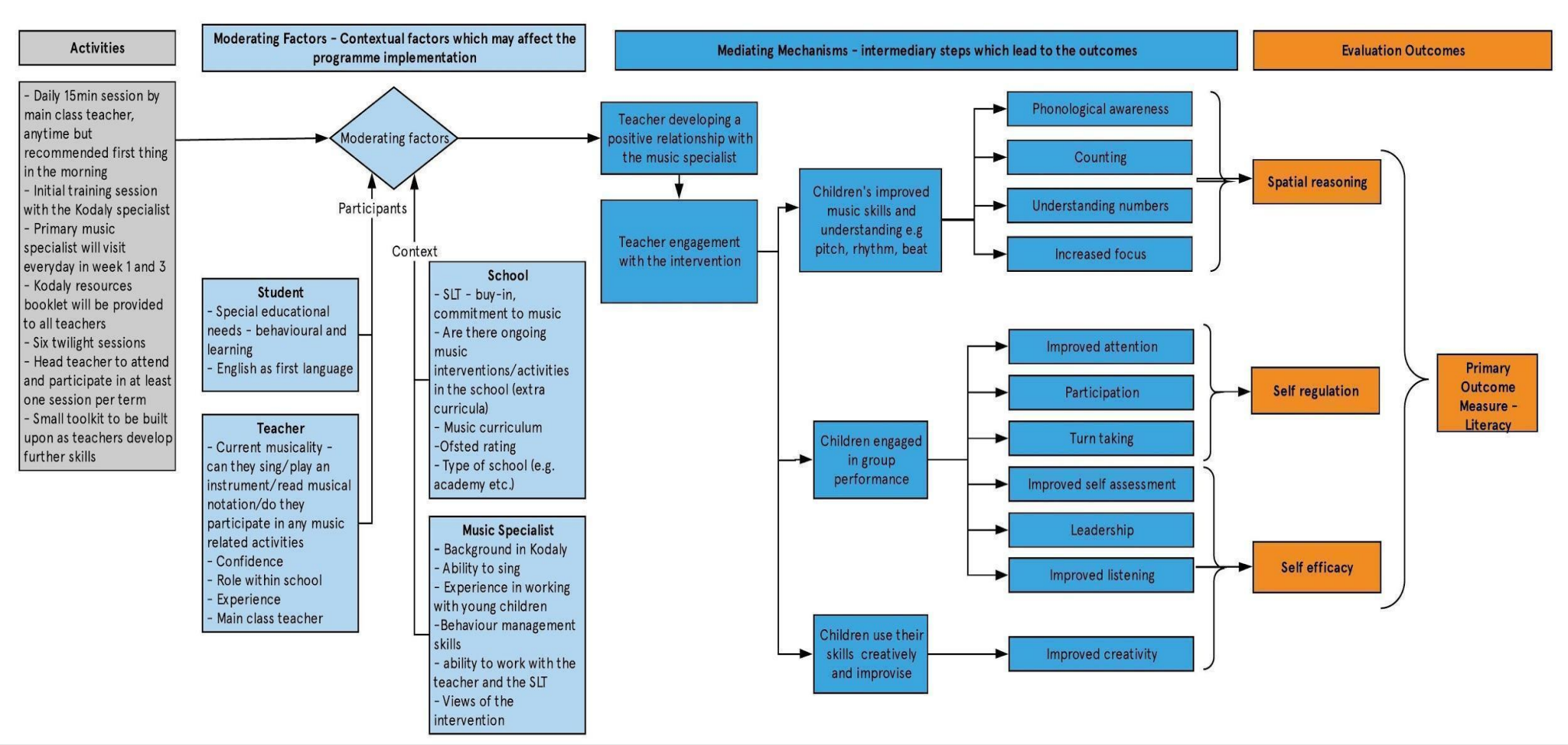
[6] https://firstthingmusic.co.uk/behaviour/

[7] The final after-school training session in the summer term was a celebratory event in June and was not counted towards training compliance.

pupils' literacy, spatial reasoning, social skills and creative self-efficacy (*Outcomes*).[8] The programme aims to achieve this by enabling teachers to develop a positive relationship with the music specialist (*Mediating mechanism*), which should then increase teacher engagement with the intervention (*Mediating mechanism*). As a result, pupils should achieve improved music skills and understanding of music (*Mediating mechanism*), which are core aims of the programme. Their increased understanding of music should lead to improvements in phonological awareness, counting, understanding of numbers and focus (*Mediating mechanisms*). Ultimately, this should lead to improved spatial reasoning. Secondly, pupils also engage in group performance (*Mediating mechanism*), which should lead to improvements in social skills, such as self-regulation. Thirdly, pupils use their music skills creatively and to improvise (*Mediating mechanism*), which should lead to improvements in creative self-efficacy. Improvements in spatial reasoning, self-regulation and creative self-efficacy should be reflected in improved literacy. These outcomes are partly dependent on several *Moderating factors*, including *pupil-related factors* such as special and behavioural needs, whether they have EAL, and their musicality; *teacher-related factors*, such as the teacher's current musicality, experience, confidence and role within the school; *music specialist-related factors*, such as their background in Kodály, their singing ability, experience working with young children, and behaviour management skills; and *school-related factors*, such as buy-in from the senior leadership team (SLT), whether there are ongoing music activities in the school, and the music curriculum.

[8] We note that spatial reasoning was not assessed as part of the current evaluation, as it was part of a larger suite of arts-based learning intervention evaluations. None of the other interventions was expected to improve outcomes related to numeracy, so spatial reasoning and numeracy were not explicitly explored in this evaluation.

Figure 1: Original logic model



Evaluation challenges

The attrition rate was 28 percent, which was higher than our assumed rate of 20 percent. Attrition was due to two schools that did not respond to arrange testing and four schools that withdrew from participating, as well as pupils changing school, being absent on testing dates, and refusing testing. Testing was conducted in autumn 2019 in two schools that had previously been unresponsive to requests, in order to minimise attrition. These schools were provided with compensation to account for additional time and disruption to accommodate testing. It is possible that programme effects may have waned by this time, but on balance, it was deemed more important to minimise attrition.

The other main challenge to the evaluation was the administration of the outcome measures. Firstly, as the PIRA is a commonly used assessment, the team learnt that at least one school had already administered the version of the assessment being used. This may have introduced a practice effect among those pupils who had already taken the assessment, but would not have introduced any bias, as the exposure would be equal among treatment and control pupils within a given school. Secondly, as pupils needed to complete the assessment independently for the most part, pupils with lower literacy skills struggled to engage with the test and some were not able to complete it.[9] Finally, the PIRA is an assessment that pupils are able to identify as an assessment, which meant some pupils were more resistant to engaging with it.

The initial evaluation plan was for all Year 1 teachers to complete the social skills sub-scale of the Social skills Inventory for each of their pupils in the evaluation. The sub-scale contains 45 items, which could take a considerable amount of time to complete for a class of 25 pupils. Given concerns about attrition and the bias that might arise from that attrition, a change was made to the protocol that the evaluation team would randomly identify 10 pupils per class for each teacher to complete the scale (more details on this is provided in the *Impact evaluation* and *Methods* sections).

Evaluation objectives

The primary objective of this evaluation was to estimate the effect of participating in FTM over the course of one school year on Year 1 pupils' reading skills. The specific research questions addressed by each element of the evaluation are further outlined in the *Impact evaluation* and *IPE* sections below.

Impact evaluation

Primary research question[10]

1. Does FTM improve pupil's reading skills?

Secondary research questions

1. Does the programme improve pupils' social skills?
2. Does the programme improve pupils' perception of their ability to generate and use ideas in their work?

[9] The PIRA manual states that teachers should follow administration 'scripts strictly to ensure results are reliable and valid' (p. 14). The PIRA 1 Summer manual instructs administrators to introduce and read the first poem and read aloud the first four questions. From this point, pupils must read everything else for themselves. If pupils are confused, administrators are not to read aloud any questions (unless instructed, as per above) or help with any individual words. See p. 14 in McCarty & Ruttle (2018). *Progress in Reading Assessment. Manual (Stage 1)*. Second Edition. Hodder Education.

[10] First Thing Music Evaluation Protocol; First Thing Music SAP.

Implementation and process evaluation (IPE)

The IPE focused on addressing research questions (RQs) related to the implementation, delivery and perceived impact of the intervention. This included both overarching implementation questions that were explored across all projects that were part of the Learning about Culture research funding, and research questions specific to FTM. RQs relevant across the set of Learning about Culture projects and those specific to FTM are detailed below.

Learning about Culture overarching IPE questions

- **RQ1** In what ways was the programme implemented? What were the barriers and facilitators of delivery (fidelity), in particular:
  - **a** SLT buy-in;
  - **b** Delivery of training and resources: (i) the extent to which it is consistent across sites; and (ii) whether it appears to be effective in ensuring that teaching staff understand the aims and main features of the intervention;
  - **c** Delivery of the intervention: (i) consistent across sites; and (ii) whether it appears to facilitate children's engagement.
- **RQ2** To what extent did the schools engage with the intervention in line with the intervention aims? (Responsiveness)
- **RQ3** How was the quality of the intervention perceived by teachers, senior leaders and TAs? (Quality)
- **RQ4** To what extent is the knowledge of arts practitioners delivering the intervention integrated with the pedagogic knowledge of teaching staff involved? (Implementer support system)

FTM-specific questions

Beyond the overarching questions which were asked, additional areas that were explored were:

- **RQ5** What are the mechanisms that are taking place in the intervention and to what extent are they bringing about change? (Mechanisms)
- **RQ6** To what extent does initial teacher confidence to deliver music lessons affect implementation, and how is the training adapted to support their needs? (Quality and responsiveness)
- **RQ7** What influences teachers' willingness to engage in music, and what music expertise do the teachers possess prior to engaging with the intervention? (Implementer characteristics)
- **RQ8** To what extent can the time be created for the intervention every day? (Fidelity)
- **RQ9** To what extent does the intervention differ from the music experience of those in the control group in class and around school? (Programme differentiation)
- **RQ10** To what extent does the intervention vary across schools, and does this affect implementation? (Adaption)

Ethics and trial registration

The project's aims, methods and materials were reviewed through the processes laid out by the UCL Institute of Education research ethics committee, and approved on 20 March 2018. While the application was approved, the reviewers stressed the importance of ensuring ongoing pupil assent for participation in any evaluation activities throughout the research. As such, all research assistants (RAs) conducting assessments with pupils verbally described the activities to the pupils using a script with age-appropriate language, informed them all activities were voluntary, and gave an opportunity for pupils to decline to participate.

Schools were informed about the trial through initial information from the developer, and formally committed to participation by signing a Memorandum of Understanding (MoU). Schools provided parents with information letters about the evaluation. These documents can be found in Appendices E and F.

This trial protocol has been pre-registered at www.controlled-trials.com, and assigned an International Standard Randomised Controlled Trial Number (ISRCTN) of ISRCTN14035536.

Data protection

As part of this project, we processed pupils' and teachers' personal data. It was important that we processed this data lawfully, following the principles laid out in the Data Protection Act 1998 (DPA) until May 2018 and the General Data Protection Regulation (GDPR) thereafter (the project spanned these two periods). We explain the lawful basis below with respect to the GDPR but there are equivalent regulations in the DPA for the justifications set out below.

BIT uses Article 6(1)f of the GDPR as the lawful basis for processing personal data as part of this project. This is generally known as the 'legitimate interests' basis. BIT carried out a 'legitimate interests assessment' in support of this. The use of pupils' and teachers' personal data as part of this research are to understand the benefits to pupils of participating in this programme in terms of their academic attainment and other related benefits. This has public benefits that BIT believes are significant in terms of understanding whether this programme has the potential to benefit children in schools across England. Without processing these data it would not be possible to provide this quality of new evidence.

UCL uses Article 6(1)e of the GDPR as the lawful basis for processing personal data as part of this project. This is generally known as the 'public task' basis. UCL has reviewed current ICO guidance,[11] and has determined that this research forms part of its performance of a task in the public interest, as one of its core purposes provided for in its Charter and Statutes. This use of data has been allocated the following UCL Data Protection Registration Number: Z6364106/2017/11/69 social research.

We do not believe that any of the data we processed falls within the definition of special category data under the GDPR. This would have required an additional justification under Article 9(2) of the GDPR.

We informed pupils' parents of the proposed data processing and provided them with the opportunity to object to this (see Appendix F for copies of the letter, withdrawal form and data privacy notice). If parents objected, then the pupils' data was never passed on to us by schools. If a parent chose to withdraw their child's data before the end of the 2018/2019 school year, then it was destroyed. The data controllers are named in the privacy information provided as part of this project, and contact details were provided should they have any queries about the data we hold about them, including provision and deletion of their data.

The information provided to parents explains in clear and plain language the lawful bases for processing (although we keep the use of technical terms in the interests of keeping the language simple), the purpose of processing the data, that they can object and this will be respected, contact details of the organisation and categories of data that we will be processing.

Data will be kept until the end of the research project, including academic paper writing and dissemination (and certainly not longer than 10 years, in line with UCL's policy on data retention). When it is deleted, it will be securely destroyed.

Data have been, or will be, shared with the DfE (part of the UK Government), the Education Endowment Foundation (who funded the trial), EEF's data archive manager FFT Education and (in a form that is truly anonymised) to the UK Data Archive. Details of this sharing are included in the Data Privacy Notice (Appendix F).

[11] https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/public-task/

Project team

The FTM programme was delivered by the FTM team led by Lindsay Ibbotson (Project Developer) with support from Susan Robertson (Manager of Tees Valley Music Service) and the music specialists. Tees Valley Music Service (TVMS) were involved with the day-to-day management of the project and provided financial and in-kind support. TVMS also supported the recruitment of schools and made connections with local schools, musicians and facilities. Lucinda Geoghegan and Zoe Greenhalgh from the British Kodály Academy provided expertise in the Kodály-based approach and supported delivery of training and the core programme. The team of music specialists was comprised of Caroline Bell, Clare Gale, Naomi Haigh, Jane Harland, Rebecca Denniff and Colette Dutot, with Olivia Wallis as research assistant (RA).

The impact evaluation was led by Kimberly Bohling and Dr Matthew Barnard, with analysis conducted by Alexander Whitefield and Fabian Gunzinger at the Behavioural Insights Team (BIT) and Dr Jake Anders and Dr Nikki Shure at UCL Institute of Education. Data collection was managed by Faisa Abdi, Eleanor Collerton, Camilla Devereux, Amber Evans, Louise Jones, Alex Manby, Bridie Murphy and Juliane Wiese. Primary data collection was carried out by RAs employed by BIT and marking of those data was also carried out by RAs employed by BIT, drawn from finishing students at UCL Institute of Education. The implementation and process evaluation was led by Dr Matthew Barnard at BIT, with analysis conducted by Lauren Crouch and input from Johanna Frerichs (BIT), Prof. Dominic Wyse (UCL IPE lead), Prof. Gemma Moss and Prof. Andrew Burn at UCL Institute of Education. The evaluation design was also supported by Daniel Carr, Dr Florentyna Farghly, Dr Jessica Heal and Dr Pantelis Solomon of BIT, and Professor John Jerrim of UCL.

Methods

Trial design

Table 2: Trial design

Trial design, including number of arms		Two-arm, clustered randomised
Unit of randomisation		Classroom (Key Stage 1)
Stratification variable (if applicable)		School
Primary outcome	Variable	Reading skills
	Measure (instrument, scale, source)	Progress in Reading Assessment (PIRA), score range 0–25
Secondary outcomes	Variable(s)	**1** Social skills **2** Creative self-efficacy
	Measure(s) (instrument, scale, source)	**1** Social skills improvement system (SSIS) – Social skills sub-measure, 46 items each scored 0–3, total raw score range 0–138 **2** Writing self-efficacy measure (WSEM) – ideation sub-measure (3 questions), 3-point Likert scale; score range 3–9
Baseline for primary outcome	Variable	Planned to be: Baseline reading attainment Protocol deviation: Free school meals (FSM) status, English as an additional language (EAL) status, class FSM composition, class EAL composition
	Measure (instrument, scale, source)	Planned to be: Early Years Foundation Stage Profile (EYFSP) aggregate score (range 4–12) for four learning goals: • understanding (FSP_COM_G02) • speaking (FSP_COM_G03)

		• reading (FSP_LIT_G09) • writing (FSP_LIT_G10 Protocol deviation: 0/1 indicator of current FSM eligibility status, 0/1 indicator of EAL status, proportion of a pupil's class reported as FSM eligible [range 0–1], proportion of a pupil's class reported as EAL [range 0–1] (all derived from school reports collected ahead of randomisation).[12]
Baseline for secondary outcomes	Variable	Planned to be: • Baseline social skills • Baseline creative self-efficacy Protocol deviation: FSM status, EAL status, class FSM composition, class EAL composition
	Measure (instrument, scale, source)	Planned to be: • EYFSP aggregate score (range 3–9) for three learning goals: • self-confidence and awareness (FSP_PSE_G06) • managing feelings and behaviour (FSP_PSE_G07) • making relationships (FSP_PSE_G08) • EYFSP aggregate score (range 2–6) for two learning goals: • exploring and using media and materials (FSP_EXP_G16) • being imaginative (FSP_EXP_G17) Protocol deviation: 0/1 indicator of current FSM eligibility status, 0/1 indicator of EAL status, proportion of a pupil's class that reported FSM eligible [range 0–1], proportion of a pupil's class that reported EAL [range 0–1] (All derived from school reports collected ahead of randomisation).

The evaluation was designed as a two-arm classroom-level clustered randomised controlled efficacy trial involving 64 primary schools with a total of 122 classes.[13] 61 classes were randomly assigned to receive the intervention and 61

[12] The reasons for the protocol deviations described in this table are provided in the *Background* section. The rationale for the choices made are detailed in the *Outcome measures* and *Statistical analysis* sections.

[13] The trial was originally planned with 65 schools and 123 classes. One school with one class of 36 pupils withdrew before randomisation, but the withdrawal was not communicated to the evaluators before randomisation. This school was randomised into control, which is why there were 62 control classes in the SAP and the trial protocol.

classes were randomly assigned to be in the control group and did not receive the intervention.[14] Where schools had two or three Year 1 classrooms (*n* = 46), all had at least one treated classroom and at least one control classroom. Where schools had only one Year 1 classroom (*n* = 18), they were allocated to either treatment or control. Condition assignment is detailed further in the *Randomisation* section.

It was acknowledged there was some risk of contamination with classroom-level randomisation; however, that risk was judged to be low. The programme relied upon specialists providing direct training and support to teachers allocated to the treatment arm, so the main risk for contamination was treatment group teachers sharing resources and tips with control group teachers. It was stressed to participating schools that sharing should not happen with teachers in the control group. Further, given that the programme required daily delivery, it was judged unlikely that control teachers would modify their daily teaching practice without the formalised training and support that is part of the FTM programme.

The primary outcome measure of the evaluation was the impact of the programme on reading attainment, measured by the Progress in Reading Assessment (PIRA) by Rising Stars. There are two secondary outcomes: the programme's effect on social skills, as measured by the Social skills sub-scale of the Social skills improvement system (SSIS), and its effect on creative self-efficacy, as measured by the ideation sub-measure of the writing self-efficacy measure (WSEM).

Participant selection

Study participants were Year 1 pupils. Schools were recruited into the study on the basis that they:

- were located in a region where more than ten schools were interested in participating in the programme, which was the case only in the North East region of England;[15]
- had discussed participation with FTM and signed an MoU detailing the conditions of participation (opt-out process, pupil data provision and endline assessment, participation in IPE activities, etc.).

Approximately 250 schools in the Teesside area, Sheffield area and other parts of the North East were approached. The recruitment strategy started with the FTM team writing to all in the list of 250 schools that appeared to meet eligibility criteria. The team then followed up with phone calls. Additionally, they held a series of four 'Recruitment / Launch' events in the four local authority areas that TVMS covers, and TVMS gave the project space at their 'Big Sing / Snappy Christmas' events in the lead-up to Christmas 2017.

Strong preference was given to two-form entry schools, as this would result in one treated and one untreated cluster (class) per school. However, the project team was unable to recruit the required number of two-form schools, so other schools (e.g., single and three-form entry) were included after discussion with the evaluation team and the EEF. Some interested schools were deemed ineligible due to location outside the North East, or having too few pupils in Year 1. Ultimately, 64 schools took part in the trial. Of these schools, 18 have one form entry, 35 have two forms and 11 have three forms.

[14] The data set includes only 60 treatment classes and 121 classes in total. This is because one school initially provided us with an incorrect roster for one of their classes, which we randomised into treatment. The school informed us of the mistake after randomisation, but did not provide us with a roster for the correct class because, in the meantime, they had decided to stop the implementation of the programme due to capacity constraints. We thus had to drop the incorrect class from the sample and were unable to replace it. Because the correct class would have been randomised into treatment in lieu of the incorrect one, had the school provided the correct information, and because we were unable to collect outcome data for the class, we treat this as a case of attrition.

[15] This requirement was necessary to make the recruiting of a music practitioner cost effective.

Schools with an average or above average share (14.1 percent[16]) of FSM children also received priority in recruitment. In the final sample, 63 classes have an above-average share of FSM children. As this trial was delivered to the whole class, there were no pupil-level eligibility requirements.

The final randomised sample consisted of 60 treatment classes and 61 control classes, with a total of 1492 and 1512 pupils, respectively. Across the treatment arms there were 121 classes with 3004 pupils.

Outcome measures

Baseline measures

Baseline measures for this research were planned to be drawn from the DfE's NPD. All participating schools were asked to provide personal information about participating pupils that would allow a reliable link to be achieved, based on current guidance from the DfE and balancing this against personal data minimisation requirements set out in data protection legislation. Using this link, it was expected that we would obtain information on pupils' performance on the EYFSP. We expected to use two communications (FSP_COM_G02, FSP_COM_G03) and two literacy (FSP_LIT_G09, FSP_LIT_G10) sub-scores as baselines for the reading attainment (PIRA) outcome. We had also selected three Social skills sub-scores (FSP_PSE_G06, FSP_PSE_G07, FSP_PSE_G08) and two creativity sub-scores (FSP_EXP_G16, FSP_EXP_G17) as baselines for the social skills and creative self-efficacy outcomes, respectively.

Due to the data access issues described above, an alternative approach was taken, with its design informed by an intention to maximise the statistical power of our analysis, by increasing the precision of our treatment estimates, given the data available. As such, instead of including the planned baseline measures in the model, we substitute the available demographic information that was collected about pupils in the year prior to implementation in order to conduct randomisation (initially intended for the purposes of stratification as part of the randomisation process), specifically eligibility for FSM at the point of randomisation and whether the child has EAL. We include these in the model as predictors themselves, and also aggregated them to the class level to produce composition variables, given evidence that school-level aggregate predictors also provide explanatory power (Bloom et al., 2007). As such, we expected the additions to improve power compared to an empty model. Nevertheless, the improvement in statistical power is still likely to have fallen short of what we would have expected from including a prior attainment measure, as was planned. It is important to understand the implications of this change. First, it is important to stress that there are no expected implications for bias in our impact estimates of not having our planned baseline measures – the unbiasedness of RCT estimates derives from the randomisation, not from statistically controlling for differences at baseline. Indeed, in principle, there is no need to include any baseline measures at all in the analysis to achieve an unbiased estimate from an RCT. Inclusion of inappropriate covariates in our analysis would have the potential to introduce bias – such inappropriate covariates are ones that could have been affected by the treatment, which is why we are including pupil characteristics from prior to randomisation. The main implication of this change is a reduction in statistical precision (i.e., the uncertainty around estimates that is inherent in all evaluations is likely to be larger in this evaluation than it would have been, which is manifested as wider confidence intervals (CIs) (or, equivalently, less likely to be statistically significantly different from zero for a given size of impact estimate). Given EEF policy to report impact estimates whether or not they are statistically significant, there is an increased risk that headline positive or negative effects are just due to this uncertainty, rather than representing a true effect, which would have been the case in the presence of more explanatory power from baseline measures (and vice versa). As a result, we particularly stress the importance of statistical significance as a check on interpretation of the results in this report.

[16] www.gov.uk/government/uploads/system/uploads/attachment_data/file/650547/SFR28_2017_Main_ Text.pdf

Primary outcome

Reading attainment

A core focus of the intervention is improved music skills and understanding of music, which might improve pupils' phonological awareness, focus and self-regulation, which support the development of reading skills (see Figure 1, *Mediating Mechanisms* and *Outcomes*). This was measured using the PIRA by Rising Stars.[17] PIRA is a standardised assessment of pupils' reading attainment and profile of reading skills. It measures reading ability in the following three areas: phonics, literal comprehension and reading for meaning.

The PIRA is a standardised and well-known test, which has been used in a number of prior EEF evaluations.[18,19] It has been shown to have high test reliability (Cronbach's alpha above 0.9), face validity (it is written to follow the national curriculum guidelines), and concurrent validity (showing a strong relationship with national test scores).[20] Another strength of the PIRA is that tests are produced at a variety of difficulty levels, graduated by school term (e.g., 'Autumn Year 1, 'Spring Year 1' and 'Summer Year 1'). We administered the PIRA Summer Year 1 to participating pupils.

End-line PIRA assessments were conducted during May–June 2019 and October 2019 by trained RAs who were blind to trial arm assignment. Assessment marking was conducted by Alpha Plus, a marking service routinely used by the assessment publisher Rising Stars, and was done while blinded to trial arm assignment. Raw scores range from 0–25.

Secondary outcomes

Social skills

The intervention involves group performance, turn-taking and other social interaction which may improve pupils' self-regulation and related social skills (see *Mediating mechanisms* and *Outcomes* in Figure 1). We assessed this using the Social skills sub-scale[21] of the SSIS.[22] The Social skills sub-scale assesses pupils' skills across the following seven sub-scales: communication, cooperation, assertion, responsibility, empathy, engagement and self-control. SSIS is a commonly used social skills assessment for young children, it is standardised, and has been used in prior EEF evaluations.[23] We chose to use SSIS over an equally popular instrument, the Strengths and Difficulties Questionnaire (SDQ) because SSIS is more thorough and in depth than the SDQ across all of the seven domains listed above. We also felt that the SSIS was a better tool to assess the types of behaviour we would expect to change as a result of participating in the programme (e.g., communication, engagement), whereas the SDQ was more oriented to identify problematic behaviours (e.g., conduct problems, peer relationship problems). The SSIS is designed to allow for

[17] This measure is not available publicly for commercial reasons; see https://www.risingstars-uk.com/Series/Rising-Stars-Pira-Tests

[18] McNally, S. (2016). *Evaluation protocol: An evaluation of teaching assistant-based small group support for literacy*. London, United Kingdom: Education Endowment Foundation. Retrieved from https://v1.educationendowmentfoundation.org.uk/uploads/pdf/Digital_-_Small_Group_Support_for_Literacy.pdf. 13

[19] McNally, S., Ruiz-Valenzuela, J., & Rolfe, H. (2016). *ABRA: Online reading support*. London, United Kingdom: Education Endowment Foundation. Retrieved from https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/EEF_Project_Report_A BRA.pdf

[20] McCarty, C. & K. Ruttle. 2018. *Progress in Reading Assessment. Manual (Stage 1).* Second Edition. Hodder Education.

[21] The SSiS contains three sub-scales: Social skills, Problem behaviours and Academic competence.

[22] This measure is not available publicly for commercial reasons; see https://www.pearsonclinical.com/education/products/100000322/social-skills-improvement-system-ssisrating-scales.html

[23] Centre for Effective Education, Queen's University Belfast. (2016). *Evaluation Protocol: Zippy's Friends*. London, United Kingdom: Education Endowment Foundation. Retrieved from: https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/EEF_Project_Protocol_Character_Zippys_Friends_protocol.pdf.

triangulation with versions of the scale for parents and teachers to complete, as well as pupils. Due to the age of the pupils and logistical challenges of collecting data from parents, the SSIS was only collected from teachers. Reliability measures are high across all versions with median Cronbach's alpha values in the mid to upper 0.90s for the Social skills sub-scale (Gresham & Elliott, 2011). Test–retest reliability is also high for the teacher version of the scale (0.81).

The questionnaires were delivered to teachers electronically. The sub-scale contains 46 items, on which teachers rate the frequency with which they observe the pupil demonstrating the behaviour; the frequency rating is then translated into point scores (Never = 0, Seldom = 1, Often = 2, Always = 3). Aggregate raw scores range from 0–138.

While it was originally planned to collect SSIS scores for all pupils, it was ultimately decided to collect them for a randomly selected subset of 10 pupils per class.[24] The reason for this was the following: the team estimated that completing a 46-item survey for 30 pupils may take teachers 2.5 hours. There were concerns that teachers might not complete the survey for all pupils in their class, and instead complete it for a non-random sub-sample (e.g., the pupils with the best behaviour), which would have biased our results. The evaluators believed that asking teachers to complete the survey for a relatively small sub-sample of their pupils improved the chance that they completed all requested surveys and – given that the research team selected pupils randomly – provided better assurance that the data would not be biased.

Creative self-efficacy

As highlighted in the logic model, the impact of the intervention on literacy may have an effect on pupils' creativity and self-efficacy (see *Mediating mechanisms* in Figure 1). For this reason, we considered pupils' self-perception of their ability to both generate and use ideas in their schoolwork (i.e., creative self-efficacy) as a secondary outcome measure. To measure this, we used an adapted version of the ideation sub-measure of the writing self-efficacy measure (WSEM) proposed by Bruning et al. (2013), with significant simplification of language to make it appropriate for this age group (the original measure was designed for secondary school pupils).[25] These adaptations were based on consultation with experts in primary literacy pedagogy at UCL.

This approach was taken to provide some scope for comparison with other trials being conducted at the same time as part of the Learning about Culture evaluations (evaluation of First Thing Music, Young Journalist Academy, Power of Pictures and Craft of Writing), in which we also examined the ideation sub-scale as part of the wider measure of writing self-efficacy. The creative self-efficacy measure was captured using three, 3-category Likert-scale items (hence, abbreviated to WSEM3; see Appendix G for the adapted scale). Each item was scored with a range of 1–3 points. Aggregate raw scores range from 3–9. The items were asked by RAs who were blinded to trial arm assignment, after completion of the PIRA assessment.

Sample size

Sample size calculations were used to determine the total number of schools and pupils required to run the trial with a suitably small minimum detectable effect size suitable (MDES). At the protocol stage, we deemed a suitable MDES to be a Cohen's *d* of 0.2. They were based on the assumptions below, with reference to the primary outcome measure, reading attainment.

1. **Randomisation was at the classroom level, stratified by school**. This was performed as specified in the *Randomisation* section.
2. **Number of arms:** There were two trial arms (treatment and control).
3. **Intra-cluster correlation (ICC):** The ICC was estimated to be 0.19. Estimating ICC values for class-level randomisations is difficult as there is less guidance available relative to school-level randomisation. Other EEF

[24] The proposed sample size of 10 pupils per class was based on power calculations, details of which are available in the Supplemental Appendices.

[25] See Appendix G.

trials that used class-level randomisation have found the estimated ICC values when performing sample size calculations were overly optimistic.[26,27] While a school-level ICC value for a reading outcome measure in KS1 would be estimated at 0.11 for schools in the North East, this was adjusted upwards to 0.19 to provide a margin of error commensurate with the experiences of previous EEF studies.[28]

4. **Cluster size:** On average, there are 22 pupils per class with five of these 22 pupils being FSM-eligible. Our assumptions were based on FSM eligibility data collected in the year prior to implementation.
5. **Attrition:** 20 percent of children in each school were expected to withdraw their data or be unable to participate in the collection of an endline outcome measure (attrition due to changing school, inability to complete assessment, etc.).[29]
6. **Hypotheses**:
    a. **Null hypothesis:** There is no difference in standardised PIRA scores between children who participate in the FTM intervention and those who do not.
    b. **Alternative hypothesis:** There is a difference in standardised PIRA scores between children who participate in the FTM intervention and those who do not.
7. **Minimum detectable effect size (MDES):** This specifies the minimum effect size our trial was powered to detect, in terms of a given standardised difference between two means of a continuous outcome measure. The required MDES for the full sample analysis was 0.20 SD (Cohen's *d*).
8. **Test–retest correlation**: The baseline achievement measure we planned to use was the Early Years Foundation Stage Profile (EYFSP). The only estimate for the test–retest correlation between the EYFSP and the reading assessment PIRA was 0.61. This was based on an unpublished analysis from the Fisher Family Trust (FFT) conducted at the end of Year 1 for a previous EEF trial (ABRA: Online Reading Support).[30] At the analysis stage, as discussed above, we were unable to access the EYFSP baseline measures as planned. We have substituted these measures with FSM eligibility (in the past year) and whether the child has EAL status, but have conservatively assumed zero impact of these variables on statistical power.
9. **Alpha and power:** We assumed 80 percent statistical power and 5 percent significance level.

It is worth noting that three key assumptions were not accurate at the point of analysis, which negatively affected our statistical power (i.e., our ability to detect an effect if one exists). Firstly, as described in the *Baseline measures* section, above, we were not able to obtain baseline data, which we previously had expected to provide a test–retest correlation of 0.61. Secondly, we underestimated the ICC. We assumed an ICC of 0.19, based on data for previous EEF studies (further details provided below), but the ICC was actually 0.24. A higher ICC indicates that pupils' outcomes within a class are more similar to one another than expected; since power calculations are based on the number of independent data points in the analysis, pupils' outcomes being more similar effectively reduces the number of independent data points and it is as if we have a smaller sample size, which reduces the MDES. Finally, our attrition rate was 28 percent instead of the estimated 20 percent, which means we have fewer observations to compare.

---

[26] Foreign Language Learning in Primary Schools, a trial testing an intervention on English literacy involving Year 3 and Year 4 children, estimated an ICC of 0.05 when performing power calculations, but found it to be 0.13 when post hoc analysis was performed (https://educationendowmentfoundation.org.uk/public/files/EEF_Project_Report_FLL.pdf).

[27] Grammar for Writing, a trial testing an intervention on writing involving Year 6 children, estimated an ICC of 0.19 when performing power calculations, but found it to be 0.32 when post hoc analysis was performed (see p. 26 in https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Grammar_for_Writing_2014.pdf).

[28] This was estimated by inflating the school-level ICC value expected by 70%, as per change between expected and observed ICC values in the Grammar for Writing trial.

[29] This assumption was based on guidance published by the EEF available at the time of writing the protocol; this guidance is no longer available.

[30] https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/EEF_Project_Report_ABRA.pdf

Based on these assumptions, the sample was planned to comprise 65 primary schools with a total of 123 classes. At the randomisation stage, that implied a sample of around 3000 children. Sample size calculations at the protocol, randomisation and analysis stages are summarised in Table 3.

Table 3: Minimum detectable effect size (MDES) at different stages

		Protocol		Randomisation[31]		Analysis	
		Overall	FSM	Overall	FSM	Overall	FSM
Minimum detectable effect size (MDES)		0.20	0.23	0.19	0.23	0.29[32]	0.38
Pre-test/post-test correlations	Level 1 (pupil)	0.61	0.61	0.61	0.61	N/A	N/A
	Level 2 (class)	0.61	0.61	0.61	0.61	N/A	N/A
	Level 3 (school)	N/A	N/A	N/A	N/A	N/A	N/A
Intra-cluster correlations (ICCs)	Level 2 (class)	0.19	0.19	0.19	0.19	0.24	0.24
	Level 3 (school)	N/A	N/A	N/A	N/A	N/A	N/A
Alpha		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

[31] Randomisation prior to one school (two classes) dropping out of the process.

[32] With baseline data, the estimated MDES would decrease to 0.23 overall and 0.30 for FSM. However, this estimate is generated with assumptions about explanatory power of the baseline measure.

One-sided or two-sided?		two-sided	two-sided	two-sided	two-sided	two-sided	two-sided
Average cluster size		22	5	20	5	21	5
Number of classes	Intervention	56	56	61	61	53	43[33]
	Control	56	56	62	62	51	43
	Total	112	112	12	123	104	86
Number of pupils	Intervention	1543	384	1509	343	1116	245
	Control	1534	384	1549	338	1034	206
	Total:	3068	768	3058	681	2150	451

Randomisation

Randomisation was conducted following the recruitment of schools, including the signing of MoUs and the completion of parental withdrawal forms and data collection processes. Random allocation was performed at the class level.

The randomisation was conducted according to the following steps:

1. Classes were stratified by school.
2. A random number was generated for each class within each school.
3. In the case of a two-form school, the class with the highest random number in the school was assigned to the treatment group, and the other class to the control group.
4. In the case of schools with three forms, half of the schools were randomly chosen to have one treatment and two control classes and the other half had two treatment and one control class.
5. One-form entry schools were grouped into a single stratum for the purposes of randomisation.

Randomisation was conducted by BIT staff using data analysis and statistical software Stata. The code used to carry out this randomisation is recorded in the Appendix H. To minimise potential for bias, the randomisation and analysis code was quality assured by a member of the UCL team to ensure the code was in line with protocol and there were no coding errors.

[33] Ten intervention classes and eight control classes did not have any FSM-eligible pupils in the analysis sample.

Statistical analysis

Primary analysis

All analyses were performed using R. The primary outcome was reading attainment, as measured by the raw PIRA total scores (0–25). In the evaluation protocol and SAP, we stated our intent to carry out the analyses using the following ordinary least squares (OLS) model in order to estimate the ITT impact of the intervention:

$$Y_i = \alpha + \beta_1 Treat_i + \beta_2 X_i + \beta_3 School_i + \epsilon_i,$$

where

- Y_i is the raw PIRA score for pupil i;
- $Treat_i$ is a binary indicator for the treatment assignment for pupil i (1 if the pupil is assigned to treatment; 0 if not);
- X_i is a vector of the relevant baseline attainment measured through aggregated EYFSP learning goal scores (specified in the *Outcome measures* section) for pupil i;
- $School_i$ is a vector of school fixed effects for n –1 schools;
- ϵ_i is the error term clustered at the class level (standard errors are corrected for clustering).

However, due to data access issues discussed above, we are unable to estimate this model due to the unavailability of the EYFSP scores as X_i. Instead, we estimate the following model in which X_i has been replaced with FSM eligibility, EAL status, class average FSM and class average EAL (as discussed in section on baseline measures):

$$Y_i = \alpha + \beta_1 Treat_i + \beta_2 FSM_i + \beta_3 EAL_i + \beta_4 School_i + \beta_5 ClassAvrFSM_i + \beta_6 ClassAvrEAL_i + + \epsilon_i,$$

where everything is as per the planned model, except that FSM_i is whether individual i is eligible for FSM in the current school year and, similarly, EAL_i is whether individual i is recorded as having EAL. In addition, we add the class-level FSM composition and class-level EAL composition, $ClassAvrFSM_i$ and $ClassAvrEAL_i$, to the regression.

Note that, while this model is a deviation from the evaluation protocol and SAP, it was planned and reported to the EEF ahead of analysis being carried out. This model has not been altered depending on the significance of any variables included (i.e., no variables were removed due to being statistically insignificant).

Secondary analysis

The secondary outcomes were creative self-efficacy as measured by the WSEM3, and social skills as measured by the SSIS. Analyses were carried out on raw scores, using the revised primary analysis model specified above, except replacing Y_i with the WSEM3 and SSIS scores. As discussed above, this model is a deviation from protocol and was made due to data access problems rendering the baseline EYFSP measures unavailable.

Analysis in the presence of non-compliance

As described in the trial protocol, teacher-level compliance in this trial was defined as having delivered 80 percent of possible sessions and missing no more than two training sessions (i.e., attending at least four out of six possible training sessions). Session delivery was recorded and shared with the delivery team on a half-termly basis. The delivery team kept records on training attendance. We estimated treatment effects for the primary outcome measure for compliers using a complier average causal effect (CACE) analysis. The CACE estimation used a two-stage least squares (2SLS) approach, which is based on the revised primary ITT model described above (Gerber & Green, 2012). At the first stage, we estimated a model of compliance:

$$Comply_i = \alpha + \beta_1 Treat_i + \beta_2 FSM_i + \beta_3 EAL_i + \beta_4 School_i + \beta_5 ClassAvrFSM_i + \beta_6 ClassAvrEAL_i + \mu_i,$$

where

- $Comply_i$ is a binary indicator for whether pupil i was taught by a teacher who met the minimal compliance threshold;
- $Treat_i$ is a binary indicator for the treatment assignment for pupil i;
- μ_i is the error term clustered at the class level (standard errors are corrected for clustering);
- all other variables are as defined in the revised primary analysis ITT model above.

At the second stage, we used the predicted values from $Comply_i$ to estimate a model of the outcome measure Y_i:

$$Y_i = \alpha + \beta_1 \widehat{Comply}_i + \beta_2 FSM_i + \beta_3 EAL_i + \beta_4 School_i + \beta_5 ClassAvrFSM_i + \beta_6 ClassAvrEAL_i + \varepsilon_i,$$

where

- Y_i is the relevant primary or secondary outcome measure for pupil i (specified in the *outcome measures* section);
- $\widehat{Comply}_i$ are the predicted levels of compliance with the programme from the first equation;
- ϵ_i is the error term clustered at the class level (standard errors are corrected for clustering);
- all other variables are as defined in the revised primary analysis ITT model above.

We conducted this analysis using the 'ivreg' function from the AER package in R (Kleiber & Zeileis, 2008). We note the deviation to protocol, with these models being based on the revised primary outcome ITT model, rather than the planned primary outcome ITT model, which is for the same underlying reasons of data access described above.

Missing data analysis

We describe and summarise the extent of missing data in the primary outcomes, and in the model associated with the analysis. Reasons for missing data are also described.

As an additional check, we ran a logistic regression model to predict missingness in the primary outcome (PIRA score) using all variables in the analysis model (treatment, FSM, EAL and school ID).

Missing covariates

In line with EEF guidelines, any imputation is restricted to the primary analysis and will only be carried out when more than 5 percent of the data is missing for a given variable. We first used a logistic regression to test whether this missing status could be predicted from all variables in the analysis model. Where predictability was confirmed (i.e., if the estimated coefficient on any of the explanatory variables in the model was significantly different from zero at the 5 percent significance level) we proceeded to the appropriate next step of this strategy. For situations for which the missing at random (MAR) assumption appeared to hold and any variable other than the outcome variable in the model was missing, we used all variables in the analysis model to estimate a multiple imputation (MI) model. MI was carried out using the Markov chain Monte Carlo (MCMC) method to predict the missing values prior to the analysis of treatment effects. We then estimated the treatment effect using the imputed data in the model associated with the primary analysis and compared our result with the primary analysis (conducted on complete cases only).

Analysis using the multiply imputed data set will be used as a sensitivity analysis (i.e., we will base confirmation of the effectiveness of the treatment on complete case analysis only, but assess the sensitivity of the estimate to missingness using the estimates from the multiply imputed data set). If the complete case analysis model implies effectiveness but the imputed estimate does not, we must assume that the missing data is not MAR to such an extent as to invalidate our conclusion of effectiveness, which we would state in the reporting of the evaluation.

Missing outcome data

Observations with missing outcome data were dropped from the analysis and a complete case analysis will be run.

Sub-group analyses

An analysis was conducted on the primary and secondary outcomes for the sub-group of FSM pupils. We originally planned to do this using the EVERFSM_6_P variable from the NPD, which indicates pupils who have ever been registered for FSM in the past 6 years. We planned to modify the original primary and secondary analysis models, with the addition of an interaction between treatment assignment and FSM status, to assess whether there is a significant difference in the treatment effect between FSM pupils and others:

$$Y_i = \alpha + \beta_1 Treat_i + \beta_2 FSMEver_i + \beta_3 Treat_i \times FSMEver_i + \beta_4 X_i + \beta_5 School_i + \epsilon_i,$$

where

- Y_i is the relevant primary or secondary outcome measure (specified in the *Outcome measures* section) for pupil i;
- $Treat_i$ is a binary indicator for the treatment assignment for pupil i (1 if the pupil is assigned to treatment; 0 if not);
- $FSMEver_i$ is a binary indicator for pupil i's EVERFSM_6_P status (1 if the pupil has been recorded as eligible for FSM; 0 if not);
- all other variables are as specified in the original primary and secondary analysis models.

As discussed above, the inability to access data required us to estimate the following revised model:

$$Y_i = \alpha + \beta_1 Treat_i + \beta_2 FSM_i + \beta_3 Treat_i \times FSM_i + \beta_4 EAL_i + \beta_5 School_i +$$

$$\beta_7 ClassAvrFSM_i + \beta_8 ClassAvrEAL_i + \epsilon_i,$$

where FSM_i is whether individual i is eligible for FSM in the previous school year (information provided by the schools). This model is identical to the revised primary analysis ITT model, with the addition of an interaction between $Treat_i$ and FSM_i. Investigation with the DfE suggests that the small difference in definition of FSM is unlikely to be material as no cleaning of the data submitted by schools is carried out before it is made available in the NPD. Given that pupils are in Year 1, we are not missing any data on historical eligibility for FSM, which is captured by the EVERFSM_6 variable. However, a limitation to this model is that this FSM variable does not capture whether pupils are eligible for FSM in the year of implementation, only in the year prior to implementation.

Additionally, in line with EEF guidelines, we estimated a separate model on the restricted sample of FSM pupils using the revised primary and secondary analysis models specified above.

Additional analyses and robustness checks

No additional analyses were pre-specified or conducted.

Estimation of effect sizes

Hedges' *g* effect size (Hedges, 1981) was calculated as follows:

$$g = J(n_1 + n_2 + 2)\frac{\underline{x_1} - \underline{x_2}}{\widehat{s*}},$$

where our conditional estimate $\underline{x_1} - \underline{x_2}$was recovered from $\beta_1$in the primary ITT analysis model;

$\widehat{s*}$was estimated from the analysis sample as follows:

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

where $n_1$is the sample size in the control group, $n_2$is the sample size in the intervention group, s_1 is the standard deviation of the control group, and $s_2$is the standard deviation of the intervention group (all estimates of standard deviation used are unconditional, in line with the EEF's analysis guidance to maximise comparability with other trials).

$J(n_1 + n_2 + 2)$was calculated as follows:

$$J(n_1 + n_2 + 2) = \frac{\Gamma\left(\frac{n_1 + n_2 + 2}{2}\right)}{\left(\sqrt{\left(\frac{n_1 + n_2 + 2}{2}\right)}\ \Gamma\left((n_1 + n_2 + 2 - 1)/2\right)\right.}.$$

If this proved intractable, we used the following approximation instead:

$$J(n_1 + n_2 + 2) \approx 1 - \frac{3}{4(n_1 + n_2) - 9}.$$

97.5 percent confidence intervals of the effect size (for primary analyses which required Bonferroni corrections; 95 percent confidence intervals for all other analyses) were estimated by inputting the upper and lower confidence limits from the regression model into the effect size formula.

Estimation of intra-cluster correlation (ICC)

The ICC of the primary and secondary outcome measures was estimated at the class level by estimating an empty variance components model, as follows:

$$Y_{ij} = \alpha + \gamma_j + \epsilon_{ij},$$

where

- $Y_{ij}$is the relevant primary or secondary outcome measure for pupil i in class j;
- γ_j is the school-level random effect;
- $\epsilon_{ij}$is the individual error term.

The school-level random effect was assumed to be normally distributed and uncorrelated with the individual-level errors. An empty variance components model is used to facilitate comparability between trials (in line with EEF guidance). The ICC itself was estimated from this model using the following equation:

$$\rho = \frac{var(\gamma_i)}{}\left(var(\gamma_i) + var(\epsilon_i)\right).$$

In the SAP, we also intended to estimate the ICC of the planned baseline measure. In a deviation from this plan for reasons of data access, as discussed above, this analysis was not conducted.

Longitudinal analysis

No longitudinal analyses were planned but we propose that, given the paucity of evidence around arts / creativity, the EEF specifically commits to following up this trial cohort to understand if there are long-term differences.

Implementation and process evaluation (IPE) methods

As part of the mixed method design of this evaluation, an IPE was conducted to complement the findings from the impact evaluation. The IPE involved members of the team with expertise and knowledge of the arts and education, which they fed into the design, conduct and analysis of the IPE. This section describes the IPE aims, sampling, data collection and analysis methods used.

IPE aims and approach

The purpose of the IPE was to understand how teachers implemented the FTM intervention in the classroom, and, in particular, to better understand the barriers and facilitators to implementation and delivering the intervention with fidelity. The IPE was also used to understand how the teaching of music through FTM compared to teachers' usual practice, and to determine the cost to schools of delivering the intervention.

Multiple sources of data were triangulated to address the IPE questions. The primary approach to IPE data collection consisted of collecting case study data from six purposively selected schools in the intervention arm of the study. Each case study involved the following methods:

1. a semi-structured interview with the teacher involved delivering the intervention;
2. a semi-structured interview with a member of the SLT;
3. informal interviews with children;
4. observations of the interviewed teacher delivering FTM.

A survey was also sent to all intervention and control group classroom teachers, and administrative data were collected from schools by the delivery team. In addition, at outset (September 2018) and midpoint (January 2019) training sessions were observed. Intervention manuals and guidance were also reviewed to inform interviews and data analysis.

The research questions and the data collection methods used to address them are shown in Table 4.

Table 4: IPE methods overview

Data collection method (case study)	Data analysis method	Participant groups	Target number of participants per case study	Actual number of participants/ activities per case study	Total number of participants/ activities	Research questions addressed
Semi-structured interviews	Framework approach	SLT	1	1	6	RQ1 (Fidelity), RQ2 (Responsiveness), RQ3 (Quality), RQ4 (Support systems), RQ5 (Mechanisms), RQ6 (Quality and responsiveness), RQ7 (Implementer characteristics), RQ9 (Programme differentiation), RQ10 (Adaption)
	Framework approach	Teachers	1	1	6 (plus one teaching assistant)	RQ1 (Fidelity), RQ2 (Responsiveness), RQ3 (Quality), RQ4 (Support systems), RQ5 (Mechanisms), RQ6 (Quality and responsiveness), RQ7 (Implementer characteristics), RQ8 (Fidelity), RQ9 (Programme differentiation), RQ10 (Adaption)
	Framework approach	Pupils	Minimum 2	4–17	60	RQ1 (Fidelity), RQ2 (Responsiveness), RQ5 (Mechanisms), RQ10 (Adaption)
Observations	Framework approach	Pupils, Teachers	1	1	6	RQ1 (Fidelity), RQ4 (Support systems), RQ6 (Quality and responsiveness), RQ7 (Implementer characteristics),

Observations	Framework approach	Trainers, Teachers	1 observation of initial training, 1 observation of midpoint training	1 observation of initial training, 1 observation of midpoint training		RQ1 (Fidelity),
Baseline survey	Descriptive statistics	Treatment teachers	47	19		RQ1 (Fidelity), RQ6 (Quality and responsiveness)
Follow-up surveys	Descriptive statistics	All teachers	47 treatment, 47 control	47 treatment, 49 control		RQ1 (Fidelity), RQ2 (Responsiveness) RQ3 (Quality), RQ6 (Quality and responsiveness), RQ8 (Fidelity), RQ9 (Programme differentiation), RQ10 (Adaption)
Administrative data collection	Descriptive statistics	Teacher session delivery	47 treatment teachers	39 treatment teachers		RQ2 (Responsiveness), RQ8 (Fidelity)
Cost interview	Framework approach	Teachers	4	4		Cost

Sampling and recruitment

Case studies

Six case study classes were selected using a purposive sampling approach to capture the range of Year 1 classes that had received the FTM intervention during the 2018/2019 academic year. The primary sampling criteria were:

1. level of engagement in the intervention (defined as high where the class teacher reported that, on average, they were delivering FTM sessions more than four times per week, and low, where they reported delivering FTM less than four times a week, or had not provided data – which also aligned with compliance criteria);
2. the proportion of pupils at the school receiving FSM (defined as high where the percentage was higher than 17.5 percent, the median for all FTM intervention schools, and low where the percentage was below 17.5 percent [information obtained from UK Government, 2019]).

The secondary sampling criteria were:

1. geographical location (categorised by local authority area);
2. Ofsted rating [recorded as Outstanding, Good or Requires improvement (information obtained from UK Government, 2019)].

Table 5 sets out the characteristics for the six case study classes that were recruited. Information is not provided on geographical location to preserve anonymity, but each school / class was based in a different local authority in England.

Table 5: Characteristics of case study schools

Characteristic	Number
Class engagement in intervention	
High (delivering First Thing Music (FTM) more than four times per week)	3
Low (delivering FTM less than four times per week or did not provide data)	3
Percentage of school pupils receiving free school meals (FSM)	
0–10%	2
11–20%	0
21–30%	1

31–40%	0
41–50%	3
Ofsted rating	
Outstanding	2
Good	3
Requires improvement	1

Surveys

Baseline and follow-up surveys were sent to school teachers taking part in the trial, via email. In October 2018, a baseline survey was distributed only to teachers whose classes were randomised to the intervention arm of the trial. Unfortunately, the response rate (calculated using the number of classes originally randomised to the intervention arm of the trial) was very low (31 percent), so it did not produce meaningful data. As a result, the baseline questions were incorporated into the follow-up survey, which was sent in May 2019. For the follow-up survey, a census approach was taken, meaning that all classes involved in the trial were invited to respond. Depending on which arm of the trial they were randomised to, teachers were asked to complete either an intervention or a control survey. The response rate for this survey was 77 percent. As the survey was a census and not a sample, confidence intervals are not given for response frequencies; in addition, as the classes taking part in the trial were not a representative sample of all primary school classes in the UK, it is not appropriate to generalise the survey findings beyond this group of classes (full details on the surveys can be found in Appendix L).

Data collection

Case studies

Schools with sampled classes were contacted by email, and where schools agreed to take part, a date was arranged for a researcher to visit this class. The researchers did not inform the delivery team about which classes they intended to visit. All visits took place between March and May 2019. At the visits, the following data were collected:

1. a semi-structured, audio-recorded interview with (i) the Year 1 class teacher involved in delivering the intervention, and (ii) a member of the school's SLT;
2. notes taken during discussions with pupils;
3. an observation of FTM teaching delivered by the interviewed teacher.

It was set out in the protocol that interviews would be conducted with teachers both before and after the observation. However, in the end only a single interview was conducted (generally after the observation), as it did not prove feasible for teachers to step out of class twice.

The interviews were conducted using guides that focused on exploring the following:

1. usual practice in relation to teaching music;
2. the context in which the intervention was implemented;
3. the facilitators and barriers to implementation;
4. children's engagement in the intervention; the perceived impact of the intervention;
5. the mechanisms underlying this change.

Interviewees 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. Discussions with children covered what they liked and did not like about FTM, what they thought of the FTM specialist, and how FTM compared to other music activities they were involved with. The researcher recorded notes during and immediately after each discussion. Full interview guides can be found in Appendix J.

Observations focused on:

1. the space and resources used to deliver the session;
2. its content;
3. details on the way the teacher delivered the session;
4. how children responded and engaged.

Observation notes were recorded first in field notes, then transferred to a structured proforma (which can be found in Appendix K). The observation data were used to help researchers probe effectively during the interviews and to deepen understanding of the observed practice.

Surveys

The research questions and programme logic model were used to inform the design of the baseline survey. The purpose of the baseline survey was to gather data on school and teacher buy-in to the intervention. These questions were incorporated into the follow-up survey for the treatment group due to low baseline response rates.

Data from interviews, as well as feedback from the delivery team, were used to inform the design of the follow-up surveys. The follow-up survey administered to intervention classes focused on:

1. teachers' confidence teaching music;
2. facilitators and barriers to attending training and delivering FTM sessions in their class;
3. teachers' views on the FTM specialist;
4. the perceived impact and quality of the intervention.

The control survey focused on understanding usual practice in terms of teaching music, and whether any learning had been shared between intervention and control classes from the same schools. Full details on the intervention and control group surveys can be found in Appendix L.

A link to the online survey platform SmartSurvey was sent to the key contacts at each school in June 2019. The intervention survey email stated that the survey should be completed by all Year 1 teachers whose classes were in the intervention, while the control survey email stated that the survey should be completed by all Year 1 teachers whose classes were randomised to the comparison group and had not attended FTM training or delivered the intervention to their class. Reminders were sent to schools who did not initially complete the survey. Those class teachers who did not respond to reminders were followed up with a phone call and given the opportunity to complete the survey over the phone. All data was collected by the end of July 2019.

Administrative data

Data were collected on the number of FTM sessions carried out each week during the autumn term (out of a total of five per week). Teachers were provided with half-termly logs to record their delivery of the programme. FTM collected the

completed logs at the end of each half term. In February 2019, the data that had been collected by FTM up to that point were provided to the evaluation team and used to inform case study sampling. At the end of the trial, these data were provided again, together with the number of training sessions attended by each class teacher (out of a total of six), as a measure of engagement with the intervention. Full data were available for 44 classes; data were not recorded for six schools which had not implemented the intervention, and two schools who did not return the data.

Data analysis

Case studies

Verbatim transcripts of the interviews and notes from the observations were analysed using the Framework approach (Ritchie et al., 2014). Firstly, emerging themes were identified through familiarisation with the data. The analytical framework was then created using a series of matrices in Excel, each relating to an emergent theme. The columns in each matrix represented the key sub-themes drawn from the findings, and the rows represented individual participants interviewed or classes observed.

The interview and observation data were then summarised in the appropriate cell, which meant that all data relevant to a particular theme were noted, ordered and accessible, facilitating a systematic approach to analysis that was grounded in participants' and schools' accounts. Analysis involved working through the charted data to draw out the range of schools' experiences and participants' views, and identifying similarities, differences and links between them. Thematic analysis (undertaken by looking down the theme-based columns in the framework) identified concepts and themes, and the case-based analysis (undertaken by comparing and contrasting rows in the framework), enabled links within cases to be established and cases compared and contrasted with each other.

During the analytical process a balance was maintained between deduction (using existing knowledge and the research questions to guide the analysis) and induction (allowing concepts and ways of interpreting experience to emerge from the data). As qualitative data can only be generalised in terms of range and diversity, and not in terms of prevalence, the analytical outputs focus on the nature of experiences, avoiding numerical summaries or language such as 'most' and 'majority'.

Surveys

Follow-up survey data were first cleaned by ensuring that all responses received came from schools with a class in the relevant arm of the trial. Following this, data were checked to identify cases where the number of responses received from a school to either the intervention or control survey exceeded the number of classes at the school randomised to each arm. In these instances, the most recent survey responses completed by respondents who identified as a 'teacher' were used for analysis. For instance, in a school with one class randomised to the intervention arm, if there were two intervention survey responses both from teachers, then the most recent one was kept, whereas if one response was from a member of SLT and one was from a teacher, the SLT response was removed even if this was more recent. Data from teachers were prioritised where there were more responses from a school than the number of classes randomised to that arm, because it was assumed that the class teacher would have been more involved in the delivery of the intervention (for the intervention survey) and know more about usual classroom practice (for the control survey), and therefore, their views and experiences were most relevant. Data was also removed from one school where the two respondents repeatedly responded with 'don't know' and said they had not delivered the intervention, so their views on the intervention were not relevant. As previously mentioned, due to the low response rate to the baseline survey, these data were not analysed.

Prior to cleaning the survey data set, there were 64 intervention class responses and 73 control class responses. Following cleaning, there were 48 responses from intervention classes (out of 61 classes randomised to the intervention group) and 50 responses from control classes (out of 62 classes randomised to the control group), giving a response rate for intervention classes of 79 percent and for control classes of 81 percent. Stata (version 14) was used to analyse

the data. Percentages scores are reported, where relevant, in the IPE findings section. Complete survey findings are provided in the Supplemental Appendices.

Costs

The evaluation gathered three key categories of data:

1. direct marginal costs (which will form the basis of the cost per pupil);
2. pre-requisites (which will be reported separately from the cost per pupil);
3. school staff time.

The data was gathered in two ways. Firstly, the evaluators requested from delivery partners information on how much they charged schools for delivering the intervention as part of the evaluation and how much they will charge schools in the future excluding any funding or subsidy that is associated with delivering the intervention as part of this evaluation. The latter data is used in calculating the cost per pupil; the former data is to ensure there is clarity about the precise nature of the data that is being requested and transparency of the approach.

The second mode of data collection was the use of case study interviews, as specified in the protocol. IPE interviews were used to determine whether questions about costs would be included in surveys, with the decision taking into account survey length and risk of damaging response rates. Based on low responses to the baseline survey, it was judged appropriate to omit cost-related questions so as to keep survey length down and not potentially dampen response rates. Instead, costs were further explored through interviews, which were also deemed a more appropriate method to gather detailed data, as they allow for follow-up questions to clarify responses and probe for more information.

The evaluation team felt that programme cost was best estimated by having a good sense of the range and diversity of experiences, which is facilitated by using a case study approach supported by purposive sampling (Ritchie et al., 2014). We selected case study schools from those who had good engagement with the programme, as they were more likely to give the best indication of the resources needed to implement the programme fully; including schools with little engagement was likely to artificially deflate costs. As an indicator of this, we selected schools from the pool who had completed the outcome measures (as fidelity data wasn't available at that stage).

The resources required to deliver the programme were most influenced by staff time and any related marginal costs, such as materials needed for implementation and travel and subsistence. We assumed these costs were most likely to be related to the amount a school has to spend per pupil and the nature of the local area and school population, the most relevant indicator for which is the percentage of pupils eligible for FSM. That is, school spending on the programme is likely related to school financial resources. Therefore, for each trial we purposively selected one school in each of the following four categories (see Table 6) in order to capture the range and diversity of costs to implement the programme.

Table 6: Categories and category definitions based on which schools were selected

Category	Category definition
#1	School with percentage of FSM in top half of participating schools and spend per pupil in top half
#2	School with percentage of FSM in top half of participating schools and spend per pupil in bottom half
#3	School with percentage of FSM in bottom half of participating schools and spend per pupil in top half

#4	School with percentage of FSM in bottom half of participating schools and spend per pupil in bottom half

The case study data was collected by RAs employed and trained by BIT. The RAs conducted interviews with teachers via telephone, using a structured interview guide designed by BIT for this purpose. RAs estimated it took approximately 20 minutes to complete the discussion.

Teachers were asked to report on direct costs of the intervention to the school, materials purchased, travel and subsistence, the cost of covering staff at training and the cost of any new physical materials purchased to improve the classroom environment. Teachers also reported on time spent embedding the intervention in their school, time at training, as well as time spent preparing to deliver the intervention. Staff were also asked to report on time taken to organise supply cover and the amount of supply cover.

Data from these interviews were used to calculate the financial and time costs outlined in this report.

Timeline

Table 7: Timeline

Dates	Activity	Staff responsible / leading
October 2017– March 2018	School recruitment	First Thing Music (FTM)
April–June 2018	Memoranda of Understanding (MoU) signing	FTM
April–June 2018	Distribution of information letters and withdrawal forms to parents	BIT
May–June 2018	Final date of return of withdrawal forms before schools send pupil data to evaluators	BIT
July 2018	Randomisation	BIT
September 2018–July 2019	Intervention delivery	FTM
September 2018	Observe first training	BIT
October 2018	IPE baseline survey	BIT
February 2019	Observe second training	BIT
March–April 2019	Conduct sampling for case studies	BIT
March–May 2019	Conduct case studies for IPE	BIT
May–July 2019	Endline assessments (PIRA and WSEM3) administered by research assistants (RAs) SSIS surveys completed by teachers	BIT
July 2019	Endline IPE survey	BIT
July–August 2019	Marking of PIRA endline assessments Data entry of WSEM3	Hodder, contracted by BIT BIT
July–September 2019	Collation and cleaning of outcomes and compliance data in readiness for upload to ONS SRS for linkage with DfE National Pupil Database (NPD) extract	UCL and BIT

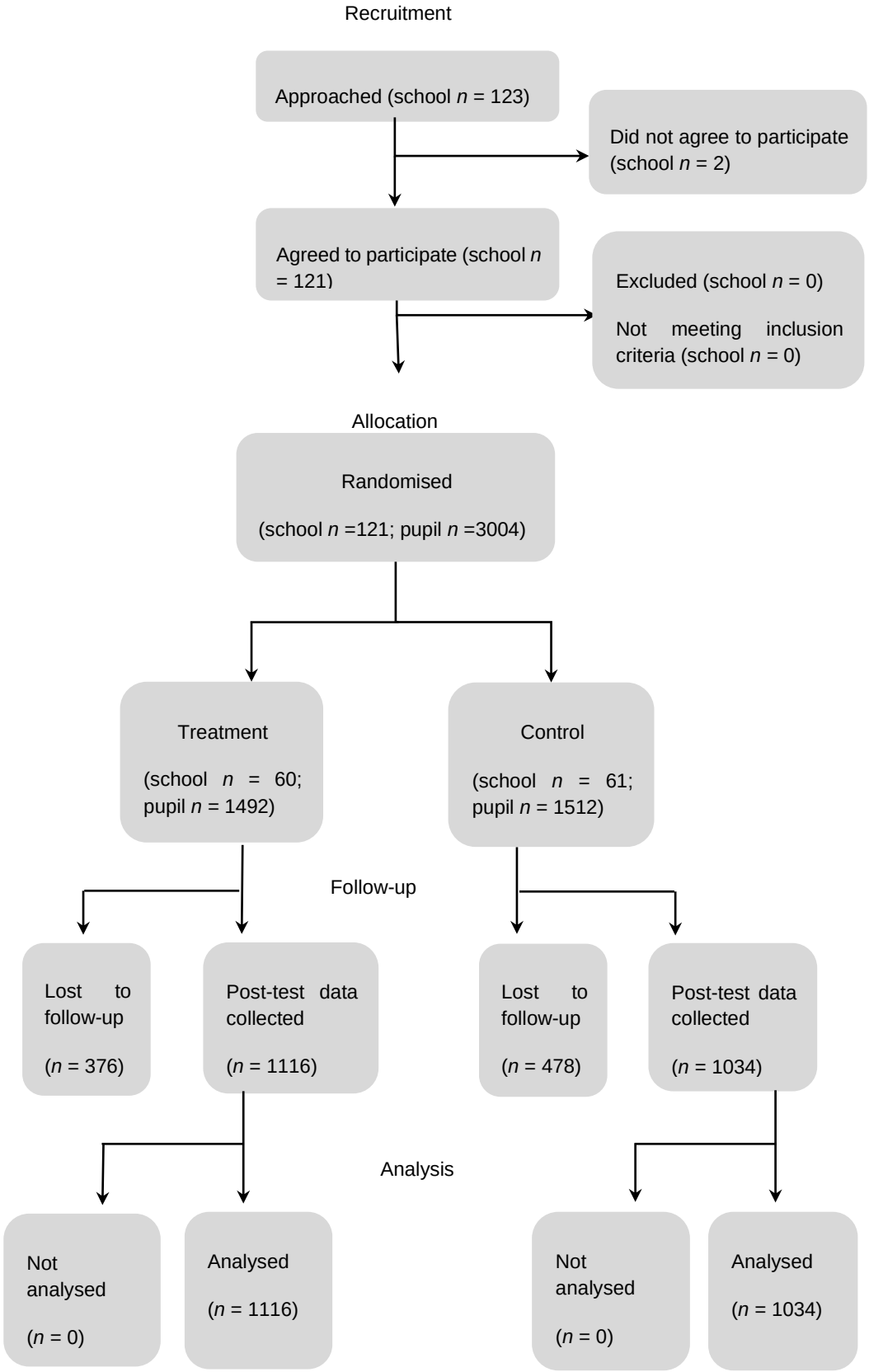
September 2019–October 2019	Cost evaluation data collection	BIT
October 2019	Testing of two schools that had been unresponsive during main testing period	BIT
January–July 2020	Project paused waiting conclusion of data sharing agreement necessary for upload of project data to the ONS SRS for linkage with NPD.	
August–November 2020	Project resumes with planned revisions as a result of not being able to achieve data sharing agreement. Impact analysis and report writing. BIT leads on the data analysis with agreed deviations from the published statistical analysis plan (SAP).	UCL and BIT

Impact evaluation

Participant flow including losses and exclusions

The flow of participants is detailed in Figure 2 for the primary outcomes. Of the approximately 250 classes that were approached, 65 schools with 123 classes initially agreed to participate in the trial and met the eligibility criteria. As detailed previously, one school (with two classes) withdrew prior to randomisation but the evaluators were not aware until outcome data collection commenced. Across these schools, 3004 pupils were randomly allocated to the intervention (1492 pupils in 60 classes) and control groups (1512 pupils in 61 classes), using a stratified class-level randomisation as described above. Figure 2 shows that there was a number of pupils who could not be reached for follow-up: 376 pupils in the treatment arm and 478 pupils in the control arm. These numbers include two schools that were not responsive when contacted about testing and four schools that withdrew from participating in outcome data collection. For two schools, data was collected in the following autumn (79 pupils total). This means that, for the PIRA outcome, a total of 1116 treatment pupils and 1034 control pupils in 121 classes were analysed for this report.

Figure 2: Participant flow diagram (2 arms) for primary outcome (PIRA score)



The MDES estimated at various points of the trial is reported in Table 3. For the PIRA outcome, this was 0.20 SD at the protocol design stage and decreased to 0.19 SD at randomisation, as more schools were randomised than expected. At the analysis stage, the MDES for the PIRA outcome increased to 0.29 due to the inability to access baseline measures as planned (discussed above), attrition and higher ICC than predicted. As pupils with FSM are a subset of all pupils, the sample size for FSM is smaller and power is substantially lower, at 0.38 at the analysis stage.

Attrition

Overall, pupil-level attrition for the primary outcome from randomisation to analysis is 28 percent. The rate is lower in the intervention group at 25 percent compared to 32 percent in the control. Of the attrited pupils, 303 (35 percent) were lost due to entire schools not taking part in testing (two schools were not responsive and four schools withdrew from testing). Of the remaining 551 attrited pupils (65 percent), this was due to pupils changing school (80 pupils), pupil absences on the day of testing (112 pupils), or pupils declining, or not available, to participate (131 pupils). We do not have information on why the test was not completed for 228 pupils.[34]

Table 8: Pupil level attrition from the trial (primary outcome)

		Intervention	Control	Total
Number of pupils	Randomised	1492	1512	3004
	Analysed	1116	1034	2150
Pupil attrition (from randomisation to analysis)	Number	376	478	854
	Percentage	25%	32%	28%

Pupil and school characteristics

Table 9 summarises the baseline pupil-level characteristics of intervention and control pupils as randomised. In general, it shows that intervention and control pupils were similar to each other, but both differed in certain ways compared to the national-level figures. The intervention group had a slightly higher proportion of FSM pupils (23.2 percent) vs. the control group (22.1 percent). Compared to the national average of 29.6 percent, the sample had a slightly lower proportion of FSM pupils. The two groups had a similar proportion of EAL pupils (intervention 10.0% vs. control 10.9%). Compared to the national average of 16.3 percent, the sample had a slightly lower proportion of EAL pupils.

Table 9: Baseline characteristics of groups as randomised

[34] Two possible reasons: 1) follow-up testing was not conducted for schools with fewer than three pupils absent on the main testing days; 2) in some cases, school timetabling meant there were fewer time periods where pupils were available for 40 minutes to complete testing, so fewer pupils were tested on the scheduled day than anticipated, and the school was not able to accommodate a second day.

Pupil-level (categorical)	National-level mean	Intervention group		Control group	
		n/*N* (missing)	Count (%)	*n*/*N* (missing)	Count (%)
Current free school meals (FSM)[35]	29.6%	1475/2987 (17)	342 (23.2%)	1512/2987 (0)	335 (22.1%)
English as additional language (EAL)	16.3%	1485/2994 (7)	149 (10.0%)	1509/2994 (3)	164 (10.9%)

Table 10 presents the analogous balance characteristics for the groups as analysed for the PIRA. Group differences between the analysed sample and national averages are broadly comparable to the sample as randomised. All analyses control for FSM, EAL and year group.

Table 10: Baseline characteristics of groups as analysed

Pupil-level (categorical)	National-level mean	Intervention group		Control group	
		n/*N* (missing)	Count (%)	*n*/*N* (missing)	Count (%)
Current free school meals (FSM)	29.6%	1116/2150 (0)	245 (22.0%)	1034/2150 (0)	206 (20.0%)
English as additional language (EAL)	16.3%	1116/2150 (0)	88 (7.9%)	1034/2150 (0)	91 (8.8%)

Table 11 summarises the baseline school-level characteristics and how they compare to national-level figures. 82.3 percent of schools in the sample were all urban, compared to the 70.8 percent of schools at the national level. There were 24 community schools, 25 academies and 13 other types of schools. 14 schools were rated by Ofsted as outstanding, 39 rated as good and six were rated as requiring improvement.

[35] Due to the inability to access NPD data (discussed above), we only have data from schools on whether pupils in the trial were currently eligible for FSM at the time of recruitment, rather than EVER FSM.

In sum, compared to the national average, the analysis sample is more urban, has fewer poor children, and is more likely to have an 'Outstanding' Ofsted rating. The treatment effect identified, may be different, if the treatment was scaled up to all schools in England.

Table 11: School characteristics as randomised[36]

School-level (categorical)	National-level mean[37]	*N* (missing)	Count (%)
Urban	70.8%	62 (0)	51 (82.3%)
Rural	29.2%	62 (0)	11 (17.7%)
Community	36.0%	62 (0)	24 (38.7%)
Academy	32.4%	62 (0)	25 (40.3%)
Other	31.6%	62 (0)	13 (21.0%)
Outstanding[38]	16.6%	59 (3)	14 (22.6%)
Good	68.1%	59 (3)	39 (62.9%)
Requires improvement	10.4%	59 (3)	6 (9.7%)

[36] Data for 62/64 schools randomised.

[37] National-level figures reflect percent FSM and EAL at KS2 (from *Compare School Performance KS2 data, 2018–2019*), given that school-level public data on KS1 FSM and EAL was not available. This comparison is limited by differences between these two stages of education.

[38] Percentages do not add up to 100 as schools with missing Ofsted ratings are included in the denominator

School-level (continuous)		*n*/*N* (missing)	Mean (SD)
Key Stage 1 (KS1) average performance	16.3	52 (10)	15.9 (1.1)
Percent free school meals (FSM) at KS2[39]	29.6	52 (10)	38.8 (21.6)
Percent English as an additional language (EAL) at KS2	16.3	52 (10)	9.4 (18.3)

Outcomes and analysis

Primary analysis

The primary outcome measure was raw PIRA score, a measure of reading attainment. Figure 3 shows the distribution of outcomes. Scores are distributed across the possible range with no observable skew; as such, we do not have any concerns about floor or ceiling effects in the analysis. Table 12 presents the results of the analysis for the primary outcome measure, PIRA scores. Further details on the calculation of effect size can be found in Appendix C. The unadjusted mean for the PIRA score in both the FTM intervention group and the control group is 11.6. After adjusting for covariates in the analysis model, the mean difference between the two groups is 0.48, which translates into a Hedges' *g* effect size of 0.07 (a higher score in the intervention group compared to the control group). This effect is not statistically significant at the 5 percent level ($p = 0.13$). This means that **being assigned to receive the FTM intervention did not have a statistically significant effect on the PIRA score**.

[39] These figures reflect percent FSM and EAL at KS2 (from Compare School Performance KS2 data, 2018-2019), given that school-level public data on KS1 FSM and EAL was not available. This comparison is limited by differences between these two stages of education.

Figure 3: Histogram of raw PIRA scores

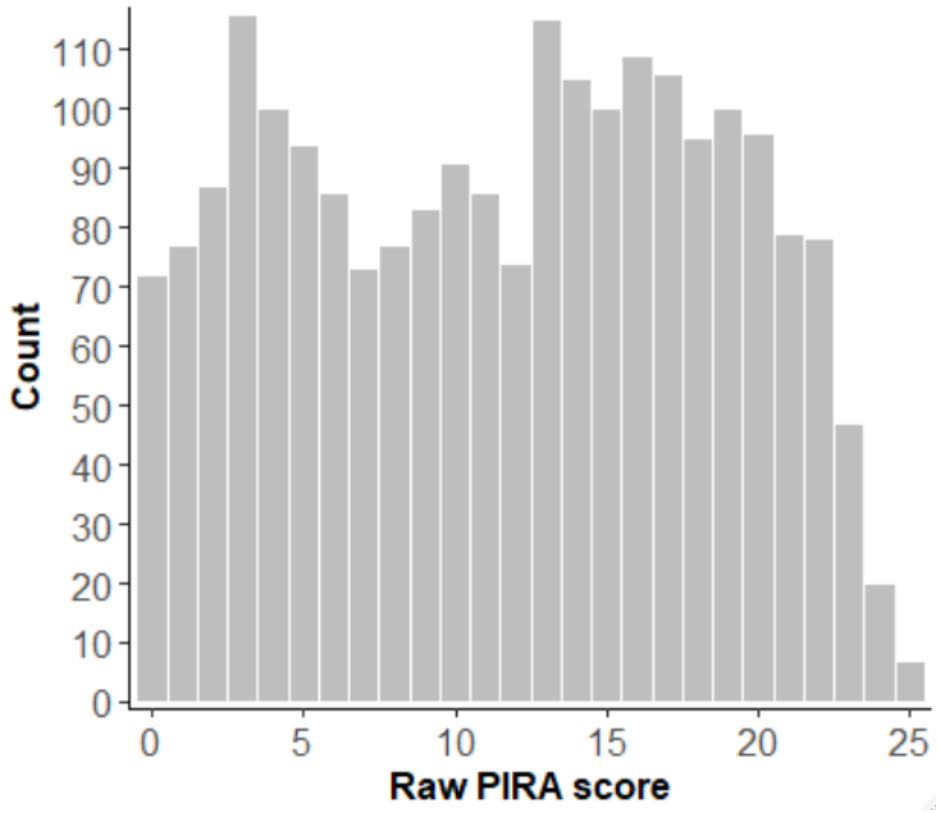


Table 12: Primary analysis

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	*n* (missing)	Mean (95% CI)	*n* (missing)	Mean (95% CI)	Total *n* (intervention; control)	Hedges' *g* (95% CI)	*p*-value
Raw PIRA score	1116 (376)	11.6 (11.2, 12.0)	1034 (478)	11.6 (11.2, 2.0)	2150 (1116; 1034)	0.07 (−0.02, 0.16)	0.13

Secondary analysis

The two secondary outcome measures were the WSEM3, a measure of creative self-efficacy, and the SSIS, a measure of social skills. Figures 4 and 5 show the histograms for these outcomes across all pupils. For both WSEM3 and SSIS scores, a higher number indicates a better score. For the SSIS measure, the distribution is skewed to the right indicating a higher proportion of teachers positively assessed pupil social skills, which is suggestive of a slight ceiling effect. The distribution of scores for WSEM3 is approximately normal, with no implications for the analysis model.

Figure 4: Histogram of SSIS scores

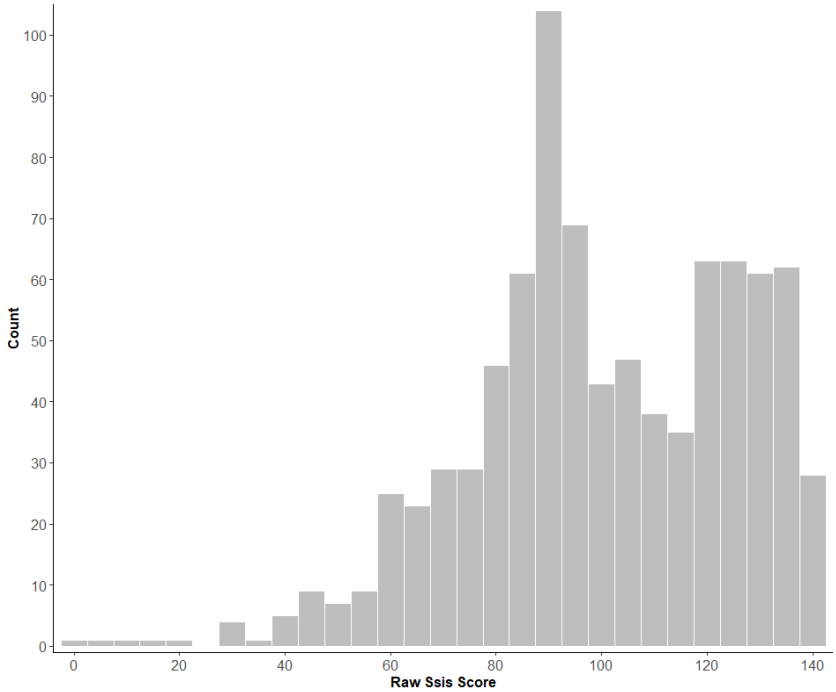


Figure 5: Histogram of WSEM 3 scores

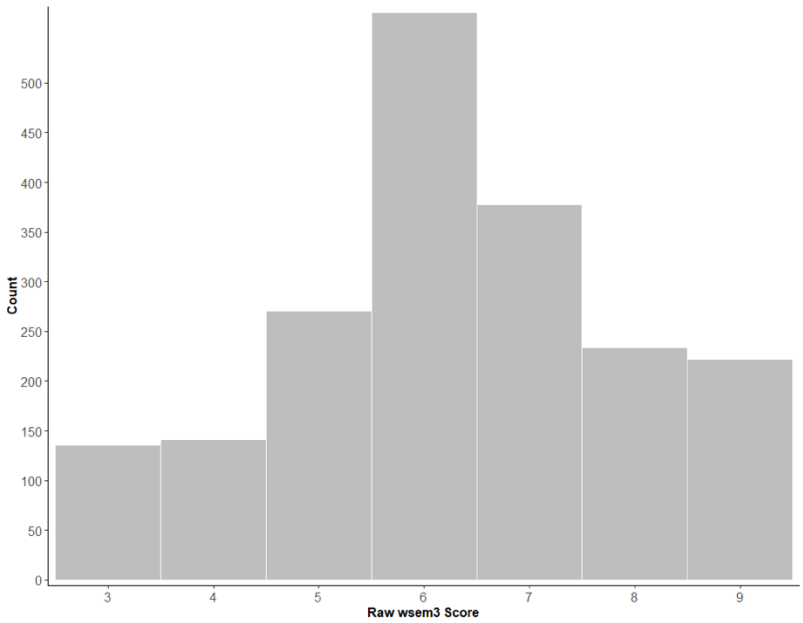


Table 13 presents the results of the analysis for the two secondary outcome measures. For the SSIS score, the unadjusted mean in the FTM intervention group is 98.4 and that in the control group is 101.7, implying social skills are, on average, better in the control group. The Hedge's *g* effect size is −0.09 (a lower score in the intervention group compared to the control group), but this difference is also not statistically significant at the 5 percent level (p = 0.27). This means that **being assigned to the FTM intervention did not have a statistically significant effect on the SSIS score.**

For the WSEM3 score, the unadjusted mean in the FTM intervention group is 6.27 and that in the control group is 6.29, implying creative self-efficacy is, on average, slightly better in the control group. The Hedge's *g* effect size for the difference between groups is −0.03 (a lower score in the intervention group compared to the control group) but this difference is not statistically significant (p = 0.51). This means that **being assigned to the FTM intervention did not have a statistically significant effect on the WSEM3 score.**

Table 13: Secondary analysis

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	*n* (missing)	Mean (95% CI)	*n* (missing)	Mean (95% CI)	Total *n* (intervention; control)	Hedges' *g* (95% CI)	*p*-value
SSIS score	433 (1059)	98.4 (96.2, 100.7)	419 (1093)	101.7 (99.2, 104.3)	852 (433, 419)	−0.09 (−0.24, 0.07)	0.27
WSEM3 Score	1025 (467)	6.27 (6.17, 6.37)	906 (606)	6.29 (6.19, 6.40)	1931 (1025, 906)	−0.03 (−0.14, 0.07)	0.51

Analysis in the presence of non-compliance

A pupil is considered to be compliant if and only if their teacher delivered at least 80 percent of possible sessions and their teacher attended at least four training sessions. Based on administrative data collected by the delivery team, compliance was low with only 40 percent of treated teachers meeting or exceeding the minimum compliance threshold. Forty-two percent of pupils in the analytical sample were in classrooms taught by teachers who reached the required level of compliance.

To estimate the impact of actually receiving, rather than simply being assigned to receive, FTM, we estimate the CACE. We estimate this using the 'ivreg' function from the AER package in R and present results in Table 14.

There were no statistically significant effects for the primary outcome. The CACE estimate is roughly double the ITT estimate. Observing a much larger effect size here is not surprising given a low level of compliance. Similar to the primary analysis, while the effect is positive, it is not significant at the 5 percent level, so we cannot conclude that FTM had an effect on those pupils whose teachers met the minimum compliance threshold.

Table 14: CACE analysis

Outcome	Intervention *N* (missing)	Control *N* (missing)	Weak instruments first-stage F test (*p*-value)	Compliance / Treatment correlation	CACE Hedge's *g* (95% CI)	CACE *p*-value
PIRA	1116 (376)	1034 (478)	F(1, 2088) = 1350.66 ($p < 0.001$)	0.42	0.17 (−0.05, 0.38)	0.13

Missing data analysis

In the SAP, we outlined a missing data strategy for the primary analysis. For missing predictor data, this strategy notes that multiple imputation would be conducted if more than 5 percent of predictor data in the model is missing. In the current case, there was very little missing predictor data (0.6 percent for FSM status and 0.3 percent for EAL status), so imputation was not necessary.

For missing outcome data, the strategy notes that a complete case analysis will be run. The extent of missing outcome data and reasons for this missingness are summarised in Table 8 (see *Attrition* section). The extent of missing data differs between groups, with the intervention group missing 23.9 percent of data for the PIRA outcome and the control group missing 31.4 percent of data. The proportion of data missing for various known reasons (i.e., pupils changing school, being absent on the day of testing, refusing testing, or the school not responding to arrange testing) is comparable between groups, apart from refusing testing, which was 6.1 percent in the treatment group and 4.6 percent in the control group.

As an additional check, a logistic regression model was run to predict missingness in the primary outcome (PIRA score) using all variables in the analysis model (treatment, FSM, EAL and school ID). We find FSM is predictive of missing PIRA scores. Results are summarised in Table 15. As specified in the SAP, we include FSM and EAL in the primary and secondary analysis. Our analysis assumes missing outcome data are ignorable. This assumption cannot be tested, though as FSM status is predictive of missingness, it is unlikely attrition is completely random (even after conditioning on FSM and EAL). Therefore, the treatment effects estimate may be biased.

Table 15: Predictors of missingness in primary outcome

Outcome	Log odds coefficient (marginal effect)	*p*-value
Free school meals (FSM)	0.35 (0.03)	0.01
English as an additional language (EAL)	0.19 (0.01)	0.38

Sub-group analyses

Sub-group analyses were conducted to identify if treatment effects differed for pupils who are, and are not, eligible for FSM. As Table 16 indicates, for all outcomes, the treatment effects are small and not statistically significant, providing little evidence of a differential effect between FSM and non-FSM pupils. For the primary outcomes, the direction and magnitude of the effect are the same as for the full sample, with higher scores in the intervention group compared to the

control group. For the SSIS and WSEM3, the direction and magnitude of the effects for FSM pupils are not significantly different to the effects for the full sample. This implies that we do not find evidence that treatment effects differ for FSM pupils compared to non-FSM pupils.

Table 16: FSM sub-group analyses

Outcome	FSM interaction term Hedges' *g* (95% CI)	Interaction term *p*-value	FSM total *n* (intervention, control)	Non-FSM treatment Hedges' *g* (95% CI)
PIRA	0.00 (−0.20, 0.19)	0.97	451 (245, 206)	0.07 (−0.03, 0.17)
SSIS	0.01 (−0.32, 0.34)	0.96	194 (106, 88)	−0.09 (−0.26, 0.08)
WSEM3	0.04 (−0.20, 0.28)	0.73	403 (220, 183)	−0.04 (−0.15, 0.07)

Estimation of ICC

We report the ICC at the class and school level for each of the primary and secondary outcomes in Table 17. It is notable that the vast majority of variance for SSIS and WSEM3, in particular, is between schools, which is relevant for other teams using this instrument as it means that studies can be planned to be smaller.

Table 17: ICC estimates for each outcome

Outcome	PIRA	SSIS	WSEM3
School-level intra-cluster correlation (ICC)	0.18	0.09	0.02
Class-level ICC	0.24	0.18	0.03

Implementation and process evaluation (IPE)

Key IPE findings

- **Usual practice:** A majority (66 percent) of survey respondents in control classes stated that their class received music lessons between once a week and once every two weeks. Treatment classes either replaced their standard music delivery with FTM or continued FTM alongside their existing music curriculum activities.
- **Adherence and compliance:** Compliance with training attendance was relatively high. The administrative data showed that 73 percent of teachers attended a minimum of four out of a possible six training sessions. However, compliance with delivery frequency was low, with only 42 percent of teachers delivering at least 80 percent of possible sessions.
- **Fidelity:** Fidelity to 'first thing' delivery was low, with only 54 percent of survey respondents indicating they delivered the sessions in the morning.
- **Implementation:** The majority of survey respondents said they felt prepared to deliver FTM (96 percent) and understood the programme's purpose (92 percent). Ninety-two percent were satisfied with the support received from the specialist, but more than half would have welcomed more support with their class.
- **Perceived impact on teachers:** In interviews, some teachers reported feeling additional pressure to deliver the programme every day. Overall though, teachers perceived FTM to have been worthwhile in terms of CPD and had enjoyed improving musically alongside the children.
- **Perceived impact on pupils:** Teachers perceived the programme to have had a positive impact on pupils' musical, social and interpersonal skills, self-esteem, self-regulation and independence. However, they also expressed some uncertainty about the intervention's ability to increase pupils' reading attainment (the primary outcome).

This section details the findings of the IPE. Data were collected via interviews, observations, discussions with pupils and surveys (see the *Methods* section for more details. The subsections are structured by theme and not research questions, in line with best practice qualitative reporting (Nowell et al., 2017); however, the research questions addressed by each subsection are outlined in Table 18.

The first subsection provides background information about the case study schools, and explores the school context in which the intervention was implemented, including teachers' perceptions of their school's programme implementation strengths and challenges at the time of delivery and the school's usual approach to teaching music. The second subsection outlines the factors that influenced the implementation and delivery of FTM, including those influencing fidelity; that is, whether it was implemented as intended by the programme developers. The third section explores engagement with the programme and the barriers and facilitators to teacher and pupil engagement. The final subsection describes the perceived outcomes and mechanisms for how FTM brought about change.

Table 18: Implementation and process evaluation (IPE) results section headings and research questions (RQs)

Section in IPE findings	Research questions
Context	RQ3. What music expertise do the teachers possess?
	RQ9. To what extent does the intervention differ from the music experience in the control group classes and school?
Implementation	RQ1. In what ways was the programme implemented? What are the barriers and facilitators of delivery? In particular: **a** SLT 'buy-in'; **b** delivery of training and resources: (i) the extent to which it is consistent across sites; and (ii) whether it appears to be effective in ensuring that teachers understand the aims and main features of the intervention; **c** delivery of the intervention: (i) consistent across sites; and (ii) whether it appears to facilitate children's engagement.
	RQ3. How was the quality of the intervention perceived by teachers, senior leadership and teaching assistants?
	RQ4. To what extent is the knowledge of arts practitioners and other practitioners integrated with the pedagogic knowledge of teachers involved?
	RQ6. To what extent does teacher confidence to deliver music lessons affect implementation, and how is the training adapted to support their needs?
	RQ8. To what extent can the time be created for the intervention every day?
	RQ10. To what extent does the intervention vary across schools and does this affect implementation?
Engagement	RQ1. To what extent did the schools engage with the intervention, in line with the intervention aims?

	RQ7. What influences teachers' willingness to engage in music?
Perceived mechanisms and impact	RQ5. What are the mechanisms that are taking place in the intervention, and to what extent are they bringing about change?

Context

This section explores the school context in which the interventions were implemented, including teachers' perceptions of their schools' programme implementation strengths and challenges at the time of delivery; the schools' usual approach to teaching music and how it compares to FTM; and teachers' attitudes towards teaching music in schools. The contextual findings documented in this section are discussed further in the report where relevant; for example, when they influence participants' engagement with or experience of the programme.

School strengths and challenges

High levels of deprivation

Of the six case study schools, four had a high number of pupils receiving FSM (defined as more than 17.5 percent of pupils in the school receiving FSM)[40]. Case study data showed that teachers in schools in deprived areas reported facing challenges due to pupils' limited language acquisition, poor mental health and lack of parental involvement. This meant some pupils may have been less prepared for starting school. For example, a teacher reported that some pupils started school without being toilet trained and had delayed speech and language development. Teachers also reported that where low-income parents worked multiple jobs, the parents could find it difficult to find time to support their children and transport them to after-school activities, or attend events (such as parents' evenings) themselves. As a result, teachers often saw that it was the same parents and pupils either engaging in out-of-school activities or not. The following quote is indicative of some of the views of the teachers:

> *'Some of our families struggle because they are on low incomes... They are really time poor because they might work in two or three jobs, so they have less time to support their children... which means that those children sometimes struggle to engage in after-school events and anything that happens outside of school.' (SLT member 06)*

Support

A nurture-based, individualised approach to supporting pupils with additional needs was described by some interviewees as a strength of some schools. Play-based learning, one-to-one support, 'Theraplay' for pupils with special educational needs (SEN), and working closely with parents of children from disadvantaged backgrounds on routine and structure, were identified as key to supporting the differing needs of pupils. Another strength included a school behaviour policy to support children with attachment and behavioural problems. Schools recognised the importance of providing a safe space for pupils to engage in opportunities that they could not access at home.

[40] As set out in the *Methods* section, FSM was defined as high where the percentage of children on FSM was 17.5% or higher, the median for all FTM intervention schools, and low where the percentage was below 17.5%.

Curriculum

The approach to the curriculum was described by class teachers and members of the SLT of the case study schools as a school strength when it encompassed an approach that was knowledge-based, person-centred and inclusive of the arts. Teachers who worked within an arts-inclusive curriculum felt that encouraging pupils to become creative, critical thinkers was a school strength. Perceptions of implementation of a knowledge-based curriculum included linking English and maths to a variety of subjects to help pupils who were not seen as strong in these subjects. A person-centred curriculum, which focused on pastoral care, safeguarding and wellbeing was regarded as a particular strength of a school in a deprived area which had a high number of disadvantaged pupils accessing FSM. On the whole, a broad curriculum that demonstrated support and inclusivity based on the needs of individual pupils was perceived to be a key school strength among interviewed teachers.

Staff expertise

Teachers identified the expertise their staff held as another strength of their school. Schools had teachers and practitioners trained specifically in de-escalation and restraint, to work with pupils who were on behaviour care plans. More generally, teachers with different specialisms worked collaboratively and shared their expertise across the school**,** and teachers particularly valued having colleagues who were music specialists, as this was perceived as rare in schools at the time of the interviews. Schools emphasised the importance of confident, committed and hard-working teachers and teaching assistants (TAs), and teachers felt such staff supported pupils to make good progress in school, despite low academic or skills-based starting points.

> *'Our children come in at a very low baseline and they leave broadly in line with expected standards across reading and maths. That's because of the sheer hard work and commitment of the staff that work here. So, the drive, perseverance and doing the best for our children is what makes us so special.'* (SLT member 03)

Usual practice

A majority of survey respondents (66 percent) in control classes stated that their class received music lessons between once a week and once every two weeks (a full breakdown of all musical activities pupils take part in at school as standard practice can be found in the Additional Appendices. In some schools, peripatetic practitioners visited and delivered music sessions. Pupils also had the opportunity to sing in assemblies and play musical instruments (both in school and privately). However, in some cases, interviewed teachers said that there was no time in the curriculum to teach music. Teachers felt music was being pushed out of the curriculum due to the pressure for schools to focus on English and maths, subjects which they felt that their school's performance is judged on.

Teachers reported that their schools used a range of published music resource packages to deliver music lessons, including Music Express and Charanga (online and physical resource schemes for teaching the national music curriculum). A teacher also used the package Go Noodle to engage the pupils in guided dance. In comparison to First Thing Music (FTM), teachers felt that Music Express did not provide as strong a focus on reading music or singing in tune compared to the Kodály method. Teachers reported that Charanga was suited to non-music specialists, but stated that pupils found the sessions boring compared to FTM because Charanga does not offer different songs every session and does not become more advanced over time. FTM was also described as offering 'short bursts' of music with more of a focus on expert singing technique than the schools' standard approach to music practice.

In interviews, treatment class teachers explained that FTM either replaced their standard music delivery or continued alongside their existing music curriculum activities. The majority of control survey respondents (92 percent) reported that musical learning had not been shared from the treatment class to their class, and some teachers in the case study schools said that they had deliberately not used the FTM approach with other classes or discussed it with other teachers as they did not want to affect the validity of the evaluation. However, other classes in some schools had heard about it and generally wanted to take part.

Teachers' previous experience and views of teaching music

Teaching and musical experience

The teachers interviewed, who were delivering FTM, ranged from newly qualified teachers who had very little experience in teaching music or musical experience themselves, to teachers with up to 26 years of experience across multiple schools. Teachers' level of music expertise also varied from teachers with little to no musical experience, to teachers who played instruments to high grades, sang in choirs and identified themselves as a music specialist. The range of teachers' musical skills was reflected in the survey findings, with the same number of respondents (40 percent) agreeing and disagreeing that they had a natural aptitude for music, though more than half of respondents agreed that they understand core music principles (60 percent). The ways in which a teacher's musical experience influenced their engagement with music and delivery of FTM are discussed in *Implementation fidelity* below.

Views on teaching music in school

Almost all survey respondents agreed it was important to include music in the curriculum (96 percent). Teaching music in school was deemed essential for helping pupils' confidence, pride, concentration, focus, memory, emotional wellbeing and social skills. Teachers also said they felt that pupils should be taught a range of subjects, and that skills developed while learning music (i.e., counting and learning patterns) could be transferred to other subjects, such as maths. Despite teachers being passionate about teaching music in schools, and pupils stating how much they enjoyed singing, teachers felt that there was less funding and time for the arts and humanities in general, and music in particular. As a result, participation in musical activities was anticipated to become more accessible to those of higher socio-economic status, if it could not be accessed freely in school.

> *'I think music is at risk of becoming almost a middle- and upper-class preserve for children, because learning [through] individual instruction is really expensive... So, if we can't deliver good quality music lessons in school, there are some children who will go through life listening to music, but never actually creating any music.' (SLT member 06)*

Implementation fidelity

As previously described, implementation fidelity refers to the degree to which the programme was implemented as described in the *Intervention* section of this report. Fidelity to the programme model varied across the schools. This section discusses five factors that were identified through case study and survey data as affecting how well implementation aligned with the intended delivery model:

1. teachers' attendance at training and how this affected their ability to deliver the intervention in class;
2. teachers' ability to deliver the intervention each day in the morning;
3. the availability of resources and classroom space;
4. teachers' confidence to deliver a music intervention;
5. the relationship and support received from the music specialist.

Each of these is outlined in turn.

Experience and application of training

Compliance with training attendance was relatively high. The administrative data showed that 73 percent of teachers attended a minimum of four out of a possible six trainings (four trainings was the minimum threshold for compliance). However, only a third of the teachers attended all six training sessions. Reasons why respondents did not attend all training sessions included: sickness; not being able to travel the distance required to the training; not receiving time off in lieu; being unable to cover the cost of travel or subsistence; and because of competing personal or school

commitments. Participants in interviews explained that attendance at training was facilitated by their school's SLT 'buying-in' to the programme and encouraging participation, for example, by being willing to provide cover.

> *'When she's had CPD [continuing professional development], I've arranged for cover in her classes to make sure that there is continuity for the children and that she has been able to go out and do the training process – we have literally let her run with it. I know the Head is totally on board with it as well.'* (SLT member 03)

Teachers who attended the training described it as practical, well-paced and a chance to troubleshoot problems. Teachers were provided with more information about FTM and generally enjoyed the participatory elements of the training. Observation of a CPD training day revealed that the main aims of FTM were discussed, including the importance of musical understanding for pupils' development of speech and literacy skills. The focus of this training session was, however, on the practical development of the teachers' skills, both in terms of understanding how to deliver FTM in a classroom setting and on building their own ability to sing the short exercises and act out the accompanying movements. Different techniques were provided that were useful in targeting the range of teachers' musical knowledge and skills, particularly for those who lacked music experience.

Perceived impact of training on delivery

The survey findings indicated that almost all respondents felt prepared to deliver FTM (96 percent), understood its purpose (92 percent), and that all respondents felt that they could deliver basic singing exercises. Learning the theory underpinning FTM helped teachers to understand why they were delivering the intervention, while observing how the music specialists delivered FTM helped teachers to understand how to deliver it. Teachers were also given resource packs during the CPD training, which they could refer to, as was seen in a case study observation when a teacher referred to their resource folder when they had forgotten a song.

The training left some teachers feeling that content such as learning rhythm would be too difficult for Year 1 pupils and that non-specialists might find the training more difficult to retain. In practice, one teacher who had little music experience prior to FTM did experience difficulty learning the songs. This may have been due to the topic taught during training not coinciding with the topic that had to be taught that term, which resulted in the teacher needing refreshers. One teacher was also unsure whether FTM required lesson plans, and how to introduce some of the songs.

On the whole, teachers found the application of training manageable in that they could be taught four songs, teach them to their class and then return to training to learn further songs. Some teachers, however, felt the training could have been improved by: spending less time standing around talking (because it did not provide teachers with skills to take back to the class); being able to view videos of FTM being delivered to pupils; and being made aware of the direction they could take teaching FTM in Year 2.

> *'I did get access to the online [videos], but it was more like teaching teachers how to do it. I found it easier watching her doing it with the class rather than looking at the tutor doing it with other teachers. It was more realistic I suppose.' (TA 06)*

Timetabling the morning intervention daily

Frequency of sessions

FTM was delivered with low fidelity and compliance in terms of the frequency of FTM sessions. Fidelity to the programme model requires daily delivery of the programme. As discussed in impact evaluation, teachers were required to deliver 80 percent of FTM sessions throughout the year (i.e., four times per week) to demonstrate compliance to the programme. If a teacher delivered 80–100 percent of sessions, this would equate to approximately 38–47 hours of instruction. According to logs maintained and submitted by teachers to FTM, 42 percent of teachers met this compliance threshold. The data showed that teachers implemented FTM an average of 3–4 days per week. However, in the survey, 73 percent

of respondents reported delivering FTM sessions 4–5 times a week.[41] Of the list of reasons for respondents being unable to deliver FTM daily, time being too limited in the curriculum and teachers feeling it was hard to prioritise FTM over other teaching were the most common (see Table 19 for full survey findings).

Table 19: Which of the following, if any, explain why you did not hold First Thing Music (FTM) every day?

Response	Number (%)
Time was too limited in the curriculum	22 (46%)
It was hard to prioritise FTM over other teaching	16 (33%)
I did not feel confident delivering the sessions	5 (10%)
The class were too disruptive to engage in the sessions	3 (6%)
The children did not enjoy the sessions	1 (2%)
I did not have the materials required to run the sessions	1 (2%)
Other: I had other commitments and responsibilities in school which meant I wasn't always in class to hold the lesson	1 (2%)
Other: I teach part-time	1 (2%)
Other: I had PPA (planning, preparation and assessment) on one of the days	1 (2%)

Note. The total percentage is greater than 100 percent because respondents could select more than one answer.

Teachers felt comfortable using curriculum time for FTM when they received support from their SLT to change the timetable; when they were aware of the benefits the initiative was having on pupils' learning; and when they felt confident that FTM was just as beneficial for pupils as the subject time that was being replaced.

> *'I found [timetabling the intervention] okay because they might [normally] be doing a morning activity... But, I think they gained just as much if not more from doing the singing than being at the table and doing phonics... I think it's just as beneficial to them.' (TA 06)*

[41] The session log data was used to calculate compliance as this was recorded and shared with the delivery partner on a termly basis, so is more likely to accurately reflect the frequency of programme delivery.

In some cases, teachers had been set timetables by their school, which meant that subjects such as maths and English could not be moved. Teachers also reported that they had too much to cover in a typical day, for instance, as well as teaching, they had to attend planning, preparation and assessment (PPA) periods on certain days. As a result, teachers either reduced the number of times a week they delivered FTM and/or moved the sessions to the afternoon. Teachers also stated that they would have preferred a more flexible approach so that they did not feel pressured to deliver the intervention every day.

Timing of sessions

Compared to frequency of delivery, fidelity was lower with regards to FTM being delivered in the morning. Survey findings indicated that only 54 percent of respondents reported that FTM was delivered in the morning. Teachers moved sessions to the afternoon when they found that FTM typically lasted longer than the recommended 10–15 minutes and led to them overrunning into the next lesson of the morning. Sessions could take longer when musical instruments had to be set up and put away, or they played games as part of the session. Sessions could be delivered more quickly as pupils' understanding of FTM gradually increased, but interviewed teachers who continued to find FTM difficult to timetable in the morning said they would have preferred to be allowed to deliver sessions when it suited them and their class. On the other hand, a teacher who found the session length suitable felt that teachers could deliver it in the morning if they were strict with their timings.

Availability of resources and space

Resources

Survey respondents and teachers interviewed had contradictory views about the availability of resources in their schools. While almost all (98 percent) survey respondents agreed that they had the materials required to deliver FTM, some interviewees reported how few instruments were readily available in the school because other year groups used them regularly. In some schools that were not well resourced, members of the SLT had been happy to buy instruments and instructed teachers to let them know if they needed to provide further resources. Teachers who did not have resources readily available to deliver FTM relied on support from management in their school to purchase or provide them.

Classroom space

During training, tables and chairs were kept in situ to show teachers how the music and activities can be delivered in a typical classroom environment. Nevertheless, teachers said that in a small classroom with lots of chairs and tables, the lack of space interrupted pupils' ability to move during the session. One teacher described rearranging the chairs and tables so that the pupils had more space to move to the beat of the songs, while another solution was to deliver the session outside, or during PE when more space is required.

> *'You can easily take [FTM] outside. There is one song that talks about splashing in puddles and the rain is falling... if it was raining and if there were puddles and they had wellies, you can just go out and do it... So there are lots of opportunities for taking [FTM] out, but it's perfectly fine for classroom bases as well.' (Teacher 05)*

Teacher confidence

Music knowledge and skills

Survey findings indicated that 58 percent of respondents had not delivered FTM daily, and 10 percent of those respondents ascribed it to a lack of confidence delivering the sessions. Other reasons cited by teachers included lack of time in the curriculum (45 percent of those not delivering daily), and difficulty prioritising over other teaching (33 percent). Less common reasons included pupils not enjoying the sessions (2 percent), pupils too disruptive to engage in the sessions (6 percent), or not having the necessary materials (2 percent). Teachers' lack of confidence teaching

FTM was underpinned by worries about not having specialist subject knowledge. Nevertheless, interviews and observations showed that teachers who were initially nervous about teaching music or singing went on to confidently deliver sessions in practice. Teachers reported being able to deliver sessions confidently because the sessions were short, practical and fun, which meant teachers were able to *'step in'* and deliver FTM. Survey findings also showed that almost all teachers (94 percent) felt that they could structure the sessions, and teachers explained during interviews that the sessions having a consistent structure helped them to feel confident in delivering the sessions.

Interviews with teachers and an observation of a training session indicated that training also boosted teachers' confidence and was adapted to account for low confidence in teachers, for example, through training facilitators using simple terminology, and acknowledging verbally that members of the group might feel nervous, but that getting involved in the training would help this.

Music specialist–teacher relationship

Specialist support and collaboration

The majority (92 percent) of survey respondents were satisfied with the support from the music specialist (e.g., stating that it was 'exceptional' and 'regular'). The music specialists offered support to teachers remotely by phone and email, and in-class by explaining and modelling concepts that teachers did not understand.

> *'She's brilliant. I will say to her, "I started this and I'm not quite sure how to work the rhythm or the notation into it." She will go, "Right, you do what you've done and I'll pick it up from there", and she'll model it for me, which has been really good. If I email her and say, "I'm stuck on this" she will go, "Right, I will come in next Tuesday", or she will say, "Look at the website we have videos...", and they've been really useful.' (Teacher 04)*

Specialists also provided teachers with resources, such as banks of new songs, which pupils also appreciated and stated was a reason for liking their music specialist. The ways in which respondents felt they could have been better supported by specialists included working more closely with the specialist and class as a whole, having more face-to-face contact with the specialist, and receiving more training from specialists (see Table 20). In addition, one respondent did say that they felt observed but not supported by the specialist, and another stated it was difficult to find the time to discuss feedback with the specialist.

Table 20: How could specialist support have been improved?

Response	Number (%)
Working more closely with the specialist and the class as a whole	27 (56%)
Having more face-to-face contact with the specialist	14 (29%)
Having more training from the specialist	7 (15%)
Working more closely with the specialist and individual children	5 (10%)

Having more online contact with the specialist	3 (6%)
Receiving more materials from the specialist	2 (4%)

Note. The total percentage is greater than 100 percent because respondents could select more than one answer.

Overall, it was helpful for teachers to be able to model their delivery of FTM based on what the specialists demonstrated, and in general the support received from the music specialists boosted their confidence. Teachers therefore recommended that the specialists visit weekly rather than fortnightly, particularly when supporting teachers who did not have a musical background. Teachers with previous musical experience did not feel they required as much support to deliver FTM.

Engagement

This section explores the barriers and facilitators to teacher and pupil engagement with FTM. For pupils, their peers' behaviour, individual needs, and the scheduling and content of the sessions influenced their initial and continued engagement. For teachers, their prior musical experiences, confidence and motivation to develop their skills influenced their willingness to engage.

Pupil engagement

Concentration

In some sessions, some pupils' lack of concentration disrupted their own and others' engagement with FTM, whereas in other sessions the majority of pupils remained on task for the duration. Observations revealed that pupils were most focused when they were engaging in singing exercises, but there were also examples of pupils fidgeting with objects from the teacher's desk and leaving the circle in which they were sitting for the FTM session. Teachers reported that, in their experience, boys were more likely to have difficulty engaging and maintaining concentration in the sessions than girls. When teachers had to attend to pupils fidgeting or lack of concentration, it was recorded in observation notes that it could lead to a slight disruption to the pace of the session. A pupil also explained how they found it difficult to hear their teacher if other pupils were talking during FTM.

Teachers used various behaviour management strategies to encourage pupils to engage with FTM, for example: asking pupils to lead songs themselves; making eye contact with pupils; having other members of staff present to help manage the class; and disciplining pupils with 'time out' if necessary. One teacher felt that it was necessary to have at least two staff members present to deliver FTM, and did not run sessions when another member of staff was not available to provide support. It was identified as particularly helpful for the other member of staff to be the FTM music specialist, so that the teacher could focus on settling disruptions, while the specialist focused on delivering FTM content.

> *'We've got a lot of behaviour [issues], so [the music specialist] coming in is great because I can work with those children and often take them away from the music session. When I'm on my own I have to keep stopping. So when she's in I can bring those children back and get them involved, which is really helpful as well.'* (Teacher 02)

Individual needs and attainment

Pupils with identified learning difficulties or SEN were described as being less engaged with FTM. Teachers reported that some pupils with learning difficulties found FTM content too difficult to follow, and some pupils with SEN would take themselves away from the sessions. A pupil reported that they had found an activity which involved swapping musical tasks in a pair *'tricky'*. The same pupil was observed during the session struggling to keep up with the activity, apparently

due to not understanding the instructions. Teachers also identified that pupils with autism struggled to engage with FTM, due to not understanding the purpose of the lesson. In one school, children with SEN were taught separately by the music specialist once a week. The teacher felt that this increased their engagement, because they were part of a smaller group and had undivided attention from the specialist.

Pupils who were shy or lacking in confidence or who had low self-esteem were sometimes reported to find it difficult to engage initially. For example, a teacher noticed pupils with social, emotional or mental health issues, who had not previously engaged in singing, initially felt embarrassed to sing. In contrast to the teachers who felt that it was difficult for pupils with specific needs to engage with FTM, other teachers felt that the programme was inclusive of all pupils. They said the intervention provided developmental opportunities for pupils with additional needs, such as refining their self-regulation and attention, while also challenging other pupils, for instance, to develop more advanced musical skills.

> *'[For] the children who typically present with special educational needs, it has been great at helping them to self-regulate and develop their attention. So, you could say it is great for them, but actually, more able pupils have been able to learn and show that they know how to write their own musical notation, and unpick a piece of music... there's that potential for self-challenge there as well, so I think it works for both more able and less able.'*[42] (SLT member 06)

In interviews, teachers explained that the mix of musical experience within a class could present challenges for delivering FTM and affect the engagement of pupils for whom music was a strength. For example, when a class did not progress onto new songs until the whole class was on beat, pupils who were already able to perform the songs could become frustrated by the lack of progress. On the other hand, when teachers progressed too quickly, pupils could forget the basics. One strategy for addressing this issue was to ensure pupils with lower levels of attainment were sat closer to the teacher, as it improved their engagement because the teacher could maintain pupils' eye contact and assist them with their hand movements, without needing to stop the session to provide additional support.

Routine and schedule

The scheduling of FTM could have an impact on pupils' level of engagement, particularly those for whom a routine was important. For example, pupils who were used to receiving music lessons in the afternoon were thrown when FTM was scheduled for the morning or did not happen at all. Pupils who felt FTM was too much of a change to their routine struggled to engage initially.

> *'I think in our school [pupils] really need a routine. Say if one day you have got something else going on, they get really flustered that you haven't done the music. Say if I did it in the afternoon, they get confused by changing it around.' (Teacher 03)*

Some teachers used their own judgement to decide when to deliver FTM, based on when they thought their pupils would be most responsive. They then adapted sessions to be shorter or longer, depending on how pupils were responding during the session. For example, a teacher found delivering FTM following reading comprehension in the afternoon worked well, because they felt that pupils were settled then. Other teachers felt pupils were more settled and responsive when they engaged in FTM in the morning.

[42] Although the terms 'more able' and 'less able' are not generally accepted terms in education, they are used here because they reflect the language used by the interviewee.

FTM content

FTM singing exercises and games encompassing movement were identified by teachers as generally effective in engaging pupils, particularly those pupils who also enjoyed singing at home. Some pupils described not liking FTM as much as other subjects such as maths and phonics, because they had trouble remembering the words and actions to the songs. One pupil reported that they did not like FTM when the songs were boring or difficult. Teachers highlighted that the games pupils played made FTM more interesting and helped pupils to focus, particularly at the start of sessions. Teachers reported that pupils with lower levels of attainment, in particular, were more engaged when walking around, and for a pupil with autism and sensory difficulties, interactive movement with the teacher during the games encouraged their participation.

Teacher engagement

Prior positive music experiences

Teachers' positive perceptions of music practitioners enhanced their willingness to take part in FTM. For example, Reception pupils at one of the schools had previously taken part in the FTM programme and the teacher felt that both the staff and pupils had enjoyed it and was therefore willing to take part this year as part of the research trial. In another case, a headteacher knew the professional musician delivering FTM in their schools' area and was therefore happy to be involved in the programme.

> *'One of my friends from the gym actually transpired to be the person who was doing the delivery... So, I was chatting to them about it and I know that they are incredibly credible with music, so I thought I would just give it a go and see what happens.'* (SLT member 02)

Teacher confidence and motivation

The intervention encouraged members of the schools' SLT to participate in one FTM session per term, but a lack of confidence also inhibited SLT members' engagement with FTM. For example, one teacher explained that the headteacher in their school did not want to engage in FTM because they were not confident singing in front of the class. The facilitator at the training also stated that headteachers had not yet attended an FTM session, despite being sent emails encouraging them to do so. Nevertheless, teachers who were motivated to learn and expand their musical skill set were enthusiastic about engaging in FTM. For example, a headteacher reported that newly qualified teachers were particularly keen to learn because they were in the early stages of their careers.

Mechanisms and impact

This section discusses the perceived areas of impact for teachers and pupils and the mechanisms through which this impact occurred. Teachers experienced additional pressures delivering FTM, but encountered positive impacts in terms of their confidence, skills and musical 'buy-in'. For pupils, teachers reported noticeable improvements in their musical, interpersonal and soft skills, but mixed levels of perceived impact on their behaviour and academic attainment. Some of the theorised mechanisms and outcomes from the original logic model emerged through the case study data. There were minor variations in the way this impact emerged, which are included in the updated logic model (Figure 6). The variations are discussed in more detail in the sections below. A broader discussion of the updated logic model, which includes findings from the impact evaluation. is included in the *Impact evaluation and IPE integration* section of this report.

Perceived impact on teachers

Additional pressure

Some teachers shared in interviews that expectations to deliver FTM every day added pressure on them, particularly on those who struggled with managing pupils with challenging behaviour. Introducing something new into the classroom like FTM (specifically the new songs) was challenging for some teachers because they felt drained by the level of behaviour management required to do so. On the other hand, when pupils were responding well to FTM, teachers found it difficult to bring sessions to an end.

Confidence and skill development

FTM provided teachers with access to musical training and opportunities, in some cases considerably more than other teachers in their school would receive. Some teachers reported that both the training and the regular, frequent delivery of music sessions had improved their music skills, and some said it had '*brought their love of music back'*. In addition, some teachers said that their self-belief and confidence to deliver music (specifically among teachers who initially had lacked confidence) had developed over the period they had delivered the programme. FTM challenged some teachers' perceptions that teaching music would be difficult, and enabled them to step out of their 'comfort zone' and try something new. A teacher had also accessed new non-musical opportunities in school as a result of FTM; for example, teaching Year 1 rather than Year 5 for the first time. Overall, teachers perceived FTM to have been worthwhile in terms of CPD and had enjoyed improving musically alongside the children.

Musical 'buy-in'

Teachers' perceptions that FTM was having a positive impact meant they were keen for key elements of the project to continue outside of the sessions and after it had formally finished. Ways in which teachers had continued to apply FTM learning outside of the sessions included:

- incorporating parts of FTM into their existing music sessions (such as Charanga);
- taking pupils to a sing-up day (which consists of pupils all over the UK joining together for one day to engage in singing and celebration of music);
- using rhythmic exercise during other subjects (such as maths) to get the class' attention.

Some teachers and SLT had discussed plans to apply the FTM principles in the upcoming year, with other year groups aside from Year 1, as well as with other Year 1 classes that did not get the chance to take part this year.

Perceived impact on pupils

The IPE found evidence for the mechanism of pupils improving their musical skills and understanding, as well as other skills, particularly those to do with working with their peers and self-regulation. A theme surrounding pupil behaviour also emerged in the IPE findings, but these findings were mixed in terms of whether the programme had a positive or detrimental effect on pupil behaviour. The original logic model included improved literacy skills as the primary outcome measure; however, findings from interviews with teachers indicated that they perceived a stronger link between the programme and improved maths skills.

Musical skills

The majority (94 percent) of survey respondents felt that FTM had had an impact on pupils' music skills. Teachers described ways in which they had noticed pupils' understanding of music gradually develop over time. For example, pupils were learning musical content (such as notations) which the teacher felt pupils in Years 3 and 4, who had not received FTM, would struggle with. A pupil also described writing down the new songs they were practising in class, in order to keep a record of their learning. Pupils' interest in and enthusiasm for music and creative tasks was also revealed outside of the FTM sessions. For example, pupils sought out musical instruments to play with at playtime and had

described singing the FTM to their parents (and even to their pets) outside of school. The opportunity to develop music skills was highlighted as particularly important in the context of not all pupils having access to individual music lessons, due to the cost of private tutoring, as well as limited other musical opportunities within some schools.

Other skills including relationships with peers

The majority of survey respondents felt that FTM has had a positive impact on a range of other skills, including social skills (94 percent), engagement (88 percent), creativity (85 percent) and communication (83 percent). In the interviews, teachers also commented on perceived improvements in pupils' concentration, listening skills, socialising and teamwork. A pupil also reported enjoying FTM in school more than playing music outside of school, because they could engage in the sessions with their friends. Teachers felt that friendship groups were bonding better, there was more kindness and tolerance between peers, and peers valued each other's ideas more. One teacher said this was because the FTM songs (such as *CopyCat*) put pupils on an equal footing because it gives them equally important roles in paired activities. Another teacher felt that, because listening skills had improved, pupils were listening to each other more. A PE teacher also discussed how a FTM treatment class was easier to teach than the control class because pupils were better at following instructions and more could be covered during lessons as a result.

> *'Our PE teacher has said, they are an easier class to teach than the other one, and they get more PE done because they'll sit and listen. They'll follow the instructions, which means that over the course of a lesson they will end up achieving more because they are not having to waste time on bringing everyone back together and saying this is what we need to do.' (SLT member 06)*

Interviewed teachers also felt that pupils' self-esteem, independence and self-regulation would improve as a result of FTM. Teachers anticipated pupils' self-regulation would improve because the regular engagement in the sessions was likely to embed in pupils' learning and provide them with self-regulation techniques. Teachers also stated that the repetition element of FTM had reinforced pupils' confidence because it enabled pupils to remember songs. More generally, FTM was described by teachers as a safe space to make mistakes, which they said helps to build pupils' resilience. This safe space was created by pupils being taught by a teacher who might not be an expert in terms of musical skills themselves, which showed pupils that they did not have to be perfect.

In addition, teachers felt that the practical nature of FTM facilitated an inclusive environment for pupils to succeed. This meant that less academic pupils who grasped FTM had the opportunity to experience performing well in a classroom setting, without requiring them to write. Teachers felt it was helpful for pupils with lower levels of attainment (who often struggled during a first lesson at school and felt down for the rest of the day), to start the day having succeeded in FTM. One headteacher reported that through pupils displaying improved attention and self-regulation, and therefore receiving less negative feedback from teachers, they would begin to experience an increase in their independence and self-esteem.

> *'I can imagine it will have had an impact on self-esteem and how children view themselves, because there are children who couldn't sit still and listen and now they can. Which means that, by extension, those children, once upon a time were likely to be constantly reminded which is quite a negative feeling for a child being constantly [told], "now you need to do this, now you need to do that". Those children are a little bit more self-regulating than they were before, which will be good for their self-esteem because they are not constantly being redirected all the time.' (SLT member 06)*

Behaviour

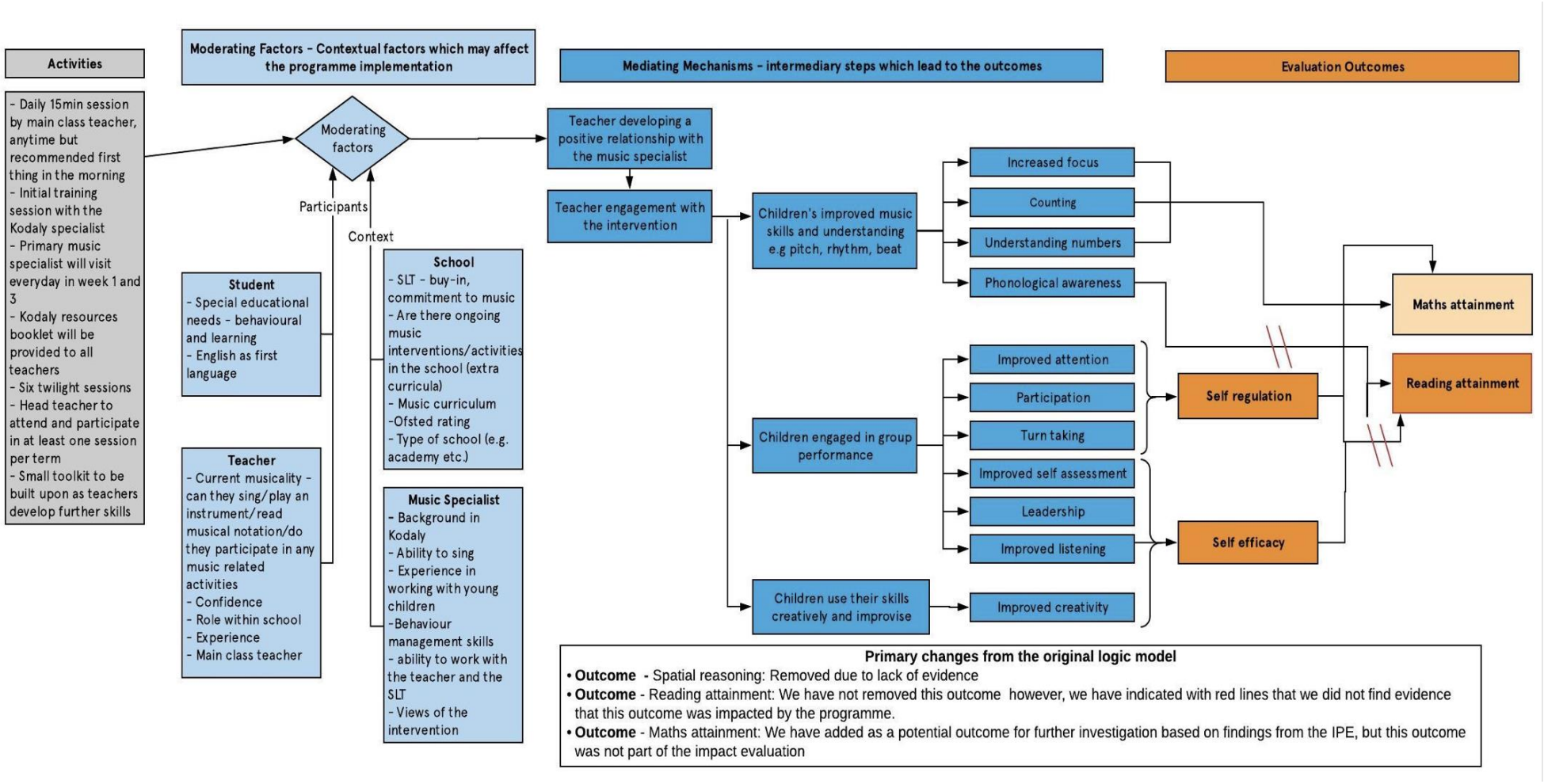
The majority (77 percent) of survey respondents felt that FTM would have a positive impact on pupils' behaviour. This was a lower proportion compared to other areas of perceived impact described above, and teachers discussed their mixed views about FTM's impact on behaviour in interviews. It was felt that afternoon FTM sessions helped to regulate pupils after lunch and, in one case, a teacher described how a class that was hyperactive when in Reception was now much more settled and pupils displayed fewer behavioural issues, which the teacher attributed to their participation in

FTM. In other schools, teachers felt that FTM had not helped with pupils' behaviour and, in some cases, had made it more difficult to manage pupils. For example, one teacher found that when the specialist visited for a session it was difficult to get the pupils re-engaged with the subject that was '*interrupted'* because their pupils did not cope well with unexpected transitions between work.

Literacy and maths

Some teachers perceived FTM to have broadened pupils' cognitive and literacy skills and vocabulary, while others felt there was no impact on academic skills, and were particularly sceptical about the potential for FTM to impact reading and writing. Such scepticism was mirrored in the survey findings, with half of the respondents feeling that FTM had neither a negative nor a positive impact on reading. Teachers did not feel that FTM would make any difference to the reading and writing skills of pupils with higher levels of attainment, and that pupils with lower levels of attainment would also not display improvements because their literacy baseline was too low to influence through this kind of intervention. Overall, teachers were aware of the research that suggests that music can contribute to improvement in literacy, but felt it was difficult to pinpoint any specific benefit associated with FTM. In contrast, it was felt that FTM had a better chance of having a positive impact on maths because of the numerical patterns and counting involved in the sessions. Some teachers did feel that FTM had led to improvements in how pupils link ideas, though others felt that pupils from more deprived backgrounds tended to see lessons in an isolated way and not make the links between subjects.

Figure 6: Updated logic model for First Thing Music (FTM)



Cost

Direct marginal costs

Delivery of the FTM intervention in the year of the evaluation cost £118,101. The majority of programme spending (£80,893) was on programme delivery, including paying for the music practitioners. Project management and administration spend was £27,618 and the remaining cost was for travel expenses to schools (£9590).

The direct cost to schools was approximately £230 per school for the year it was delivered. Programme fees for schools participating in this evaluation were £200 per school, as the fees were subsidised by the EEF. In subsequent years without subsidisation, the direct costs of programme delivery would be passed on to schools. To calculate future programme fees, we divided the total delivery cost by number of schools (*n* = 64), which would result in a school fee of £1845.

To calculate the total cost per pupil over 3 years, we assumed that the number of pupils would remain stable at 25 pupils per year, cumulatively increasing to 75 pupils treated in total by Year 3. We assumed 25 pupils as this was roughly the median number of treated pupils per school. We also assumed no subsidisation of the programme fees in Years 2 and 3. Based on these assumptions, the total cost per pupil, per year over 3 years is £52 (set out in Table 21).

Table 21: Cost of delivering First Thing Music (FTM)

Item	Type of cost	Average cost (minimum, maximum)	Total cost over 3 years	Total cost per pupil per year over 3 years
Materials and printing	Start-up cost per school	£8 (£0, £10)	£8	<£1
Travel expenses (to attend CPD)	Start-up cost per school	£22 (£0, £50)	£22	<£1
Programme fees (paid to delivery partner)	Annual cost	Year 1: £200 Years 2 and 3: £1845 per year	£3890	£52
Total		Year 1: £230 Years 2 and 3: £1845	£3920	£52

The cost breakdown is set out in Table 22.

Table 22: Cumulative costs of First Thing Music (FTM) (assuming delivery over three years)

	Cost (£)	Number of pupils	Cost per pupil	Cost rating
Year 1	£230	25	£9	£
Year 2	£2075	50	£42	£
Year 3	£3920	75	£52	£

Time costs

Training

To support the delivery of the intervention, teachers in treatment classes were asked to attend one full-day training session and six three-hour training sessions. Schools used either internal staff or external agency staff to cover this time, depending on individual schools' standard practice. All participating treatment teachers are expected to attend the training, so schools with more than one treatment class had to cover more than one absent teacher. Cost survey respondents estimated that schools staff spent an average of 45 minutes arranging supply cover over the course of the programme.

Preparation

There was additional administrative time associated with supporting the intervention. Teachers reported that in the first year they spent approximately 6 hours per school preparing for the programme (ranging from 0 to 15 hours over the course of the year.) Assuming the same level of preparation is required each year, staff time for preparation over three years would amount to approximately 18 hours. Cost survey respondents also reported spending, on average, approximately 2.5 hours over the course of the year liaising with the programme specialist via text or email, responding to emails.

Delivery

The recommended amount of staff time required for the delivery of the intervention is 15 minutes per day, five days a week. Some teachers indicated they spent 20 minutes per day delivering the intervention. On average schools reported delivering 51 hours per treatment classroom per year. The reported hours of delivery ranged from 42–58 hours per classroom per year.

Conclusion

Table 23: Key conclusions

Key conclusions
1. Pupils in the First Thing Music (FTM) intervention group made the equivalent of one month's additional progress in reading, on average, compared to pupils in the control group. This is our best estimate of impact which has a low to moderate security rating. However, as with any study, there is uncertainty around the result: the possible impact of this programme ranges from no additional progress to positive effects of three additional months of progress.
2. Pupils in the FTM intervention group achieved lower scores in an assessment of social skills than pupils in the control group at the end of the trial, though the difference in scores was small and there is some uncertainty around this result. There was no evidence from the impact evaluation that the FTM intervention had an impact on creative self-efficacy.
3. Only 40 percent of teachers in the intervention group delivered at least 80 percent of the possible music sessions and attended at least four of the six training sessions. The impact of the programme on reading attainment was higher, on average, in classrooms where this threshold for training and delivery was met.
4. Around half of the teachers (54 percent) delivered the programme 'first thing' as planned, while other teachers moved the sessions to later in the day. Timetabling difficulties, such as the sessions overrunning, were a barrier to delivery in the morning.
5. Teachers perceived the programme to have had a positive impact on pupils' musical skills, social skills, creativity and self-regulation, but more than half of those surveyed felt that the programme was unlikely to impact reading attainment.

Impact evaluation and IPE integration

Evidence to support the logic model

Results from the evaluation support the original logic model to some extent, with the predominant variance from the logic model occurring with regards to compliance to training, fidelity to intervention delivery and the impact of the intervention on the primary outcome identified.

For example, prior to implementation some schools did not attend the number of training sessions required, and, during implementation, some schools turned to delivering the sessions in the afternoon as opposed to the morning. Although a key activity and fidelity indicator was to deliver the interventions in the morning, teachers appreciated more flexibility in the timing of sessions and highlighted key benefits to delivering sessions at a time that suited their pupils best. If this intervention was tested again, we would recommend a key change to the timing of the delivery of the daily sessions in the logic model.

We found limited evidence of the use of specific toolkits that were built upon as teachers developed, headteachers' presence at sessions, and whether the music specialists' experience of working with young children or ability to sing influenced any subsequent mechanisms that brought about change in outcomes as part of the programme. Members of SLT did, however, display buy-in (as documented in the logic model) through other means, such as providing resources which facilitated the teacher's ability to deliver the interventions. We therefore recommend removing 'headteacher presence at sessions' and replace this with 'headteachers providing necessary resources and securing time in the curriculum as a key activity'.

All mediating mechanisms documented in the logic model (Figure 1) were identified in the IPE findings; however, the perceived impact of these differed slightly from the outcomes identified in the model. Notably, the impact evaluation found no evidence that FTM impacted pupil literacy skills. Survey findings mirrored this, with half of respondents indicating that they did not expect FTM to positively or negatively affect reading and writing skills. However, some participants felt that FTM would have more of a positive impact on pupils' maths attainment in comparison to reading and writing attainment, due to the numerical components involved in the FTM content. There was no perceived impact on spatial reasoning, so this element has been removed from the updated model (Figure 6); instead, there were

perceived positive impacts of the intervention on other skills, such as confidence, communication and peer relationships. However, the impact evaluation did not find any significant impact on pupils' social skills. We would therefore recommend including the further impacts that emerged from the evaluation (as documented in this report) as outcomes in the logic model, with maths attainment replacing the literacy attainment measure as the primary outcome.

Interpretation

FTM is a thoughtfully designed intervention that aims to improve music understanding and literacy and social skills through daily short music lessons. The evaluation of FTM sought to add to the existing research base on cultural learning interventions, which suggests that these interventions can go beyond aiding communication and social skills and are associated with improvements in more academically focused areas such as mathematics and literacy (Catterall, 2009; Newman et al., 2010; Catterall et al., 2012). However, the evidence on music education in particular, and its impact on pupil attainment is weaker and mixed. As such, this evaluation set out to assess the extent to which FTM could have a measurable impact on reading skills (the primary outcome), as well as writing self-efficacy and social skills (the secondary outcomes).

The outcome of the evaluation was that no statistically significant effects were found for either the primary or secondary outcomes. The measure of reading indicated that the average outcome for the intervention group was marginally better than for the control group (but not statistically significant). For writing self-efficacy and social skills, the average outcome for the intervention group was marginally worse than for the control group, but again, the difference is not statistically significant. When looking at the sub-group of pupils in receipt of FSM, the direction and magnitude of the effect are similar to those of the primary analysis and the difference was not significant.

Overall compliance with delivery and training was relatively low, with only 40 percent of teachers meeting the threshold. Administrative data indicated that only 42 percent of teachers delivered at least 80 percent of all possible sessions (i.e., at least four times per week). Instead, the data showed that, on average, teachers were delivering the programme 3–4 times per week. Further, only slightly more than half of teachers delivered the programme 'first thing', as some teachers had difficulties adjusting their morning timetable in order to accommodate the programme every day. Training compliance was much higher, with 73 percent of teachers attending at least four of the six training sessions offered. When the analysis took overall compliance into account, the observed effect on reading scores was twice as large as that observed in the primary analysis, but the difference was not significant (Primary: $g = 0.07$, 95% CI = −0.02, 0.16; CACE: $g = 0.17$, 95% CI = −0.05, 0.38). These findings should be interpreted with caution as they represent a small non-random sub-sample of the most engaged classrooms, but they may indicate that high engagement with FTM training and regular delivery are important to impacting the programme's identified outcomes.

Teachers did indicate some uncertainty about whether FTM would have an impact on pupil reading attainment; more than half of survey respondents indicated that they felt the intervention would have neither a negative nor a positive impact on reading. In contrast, some teachers felt that FTM had a better chance of having a positive impact on maths because of the numerical patterns and counting involved in the sessions. Assessing impact on literacy was prioritised over numeracy in this evaluation in order to allow for more comparisons across the five interventions evaluated as part of the Learning about Culture programme. Further research on the programme's impact on numeracy may be beneficial, and it may be worth considering follow-up analysis of KS2 maths SATs. However, it is worth noting that a recent meta-analysis of 54 experimental studies on music training with children found no evidence to support the hypothesis that music training impacts children's cognitive skills (Sala & Gobet, 2020). In fact, the authors were much starker in their interpretation about the 'far transfer' between music education and improved attainment – so much so that we believe this should restrict the EEF's openness to further testing of music-based interventions that claim to impact on attainment measures, irrespective of the result from this trial:

> *'Music training has repeatedly been claimed to positively impact children's cognitive skills and academic achievement (literacy and mathematics). This claim relies on the assumption that engaging in intellectually demanding activities fosters particular domain-general cognitive skills, or*

> *even general intelligence. … [A] considerable amount of cross-sectional evidence indicates that engagement in music has no impact on people's non-music cognitive skills or academic achievement. We conclude that researchers' optimism about the benefits of music training is empirically unjustified and stems from misinterpretation of the empirical data and, possibly, confirmation bias.' (Sala & Gobet, 2020: 1429)*

The majority of teachers did express a belief that FTM positively impacted pupil musical skills (94 percent), social skills (94 percent), engagement (88 percent), creativity (85 percent) and communication (83 percent). However, the secondary analysis on creative self-efficacy and social skills did not support this observation.

Overall, the evaluation does not indicate that FTM, as delivered in the educational context as it was when the study was carried out, has a measurable impact on either the primary or secondary outcomes. Given the low rates of implementation fidelity, it may be worth revising the intervention and undertaking further research on future iterations.

Limitations and lessons learnt

Limitations

As previously detailed in the discussion of sample size calculations, three assumptions in our initial power calculations did not hold, all of which negatively affected statistical power (i.e., the ability to detect an effect of the size anticipated if one exists). Firstly, due to issues in accessing the NPD, planned baseline measures (planned to be EYFSP scores) were not available for this evaluation. Additionally, both the ICC and attrition rates were higher than predicted for the primary outcome variable (PIRA). As such, the MDES at analysis is 0.29, compared to an assumed MDES of 0.19 at randomisation. If the pre-test/post-test correlation of 0.61 is accurate, the MDES would decrease to 0.23 with the use of baseline assessment data. Due to the data access issues, it was also not possible to obtain the FSM variable indicating whether pupils were ever eligible for FSM in the past six years (EVERFSM_6_P), and instead there was only access to information indicating whether pupils were eligible for FSM **at the time of recruitment (i.e., in the summer term of the year prior to implementation).** It is therefore possible that the evaluation had missed relevant pupils who were newly eligible for FSM while the programme was being implemented for this sub-group analysis. However, given the age of the pupils (Year 1), there was no missing historical FSM eligibility data that is captured by the EVERFSM_6 variable, as the FSM variable used in the analysis reflected their status in Reception. Based on this information, it may be worthwhile to conduct the analyses again with the pre-specified baseline outcome data.

In relation to data access issues, the planned baseline outcome measures were not accessible, so it is not possible to verify baseline equivalence between the treated and untreated pupils. However, data on FSM and EAL does not suggest that the groups are substantially different. As indicated in Table 9, the intervention group had a similar proportion of FSM pupils (23.2% intervention vs. 22.1% control) and EAL pupils (10.0% intervention vs. 10.9% control). This equivalence is consistent with randomisation having worked.

There are limitations related to the primary outcome measure, the PIRA. Administration guidelines required pupils to complete most of the assessment independently. However, some RAs administering the test observed that pupils with lower literacy skills struggled to engage with the test, and some were not able to complete it. Additionally, the PIRA is an assessment that pupils are able to identify as an assessment, which meant some pupils were more resistant to engaging with it, especially as participation was voluntary. However, any difficulties related to test administration would apply equally to both control and treatment pupils, so it is unlikely this would bias the results. It was also previously noted that some schools had already administered the PIRA, which may have introduced a practice effect among those pupils. However, as most schools had both treatment and control classrooms, this effect would have been observed equally across arms and was not likely to bias the results.

Attrition also carries with it a risk of bias (Sterne et al, 2019) and one that cannot typically be remedied analytically. In short, those schools and pupils with missing outcomes data could have affected the result from the trial in different ways

– meaning that the result is biased away from the 'true' estimate of impact. However, school withdrawals had the largest impact on pupil attrition, and at least two of the four withdrawn schools had reported that they had done little to no implementation of the programme. Given the ITT design of the evaluation, had we been able to conduct testing at more of the withdrawn schools, our impact estimates on the treated group would have taken into account more pupils who had not actually received the treatment.

Generally, the data collected from the six case study schools as part of the IPE (via either interviews or fieldwork visits) only represent the views and experiences of a subset of the larger treatment population (i.e., the selected case study classrooms). The qualitative findings are therefore not statistically representative, though the use of purposive sampling means that they should provide a good indication of the range and diversity of experiences and attitudes. Additionally, there may be some recall errors in survey responses.

Finally, the case study approach to the cost evaluation represented the range and diversity of costs encountered in implementation among highly engaged schools. Sampling was done to capture variation of spend among schools with high/low proportions of FSM pupils and per pupil spending, as these were hypothesised to correlate with costs. It is possible that the sampled schools were not representative of typical costs of full implementation of the programme. However, the programme fee was the largest cost to schools and did not vary by school. Spending ranges were provided for other direct costs, so that prospective schools could consider costs they may encounter above and beyond paying for the programme itself.

Lessons learnt

There was a risk of contamination using a classroom-level treatment assignment. As most schools contained both treated and untreated classrooms, it was feasible that teachers in the treatment group might share information with teachers in the control group. At the design stage, this risk was judged to be low (see *Trial design* section for more detail), which bore out in practice. Survey data from teachers in the control group indicated that only 3 of 50 respondents (8 percent) had had learnings about the programme shared with them. Only one of those three indicated this information had influenced their teaching practice. This finding is promising when considering utilising within school randomisation for future evaluations.

One challenge related to outcome measures is using one that is commonly used by schools. Commonly used assessments carry the benefits of being familiar to schools and having more data on reliability, validity and correlations with other common assessments. However, this can create challenges where the assessment is already being used by participating schools – both for schools who plan to test before, or after, the evaluator does. It might be worth future evaluators considering data-sharing arrangements with schools, particularly where the evaluator anticipates conducting the assessment first.

As previously mentioned, school withdrawals and non-responsiveness to requests to schedule testing contributed to attrition. In some cases, schools identified early in the year that they would not be able to implement the programme as planned (often due to staff capacity constraints). Where this was the case, it was challenging for both the delivery partner and the evaluator to convince some schools that it was worthwhile to engage with the testing process at the end of the school year, since they had not directly benefitted from being part of the evaluation as originally planned. Some schools were very committed to supporting research regardless and were able to accommodate testing; but where schools had already identified capacity constraints related to programme implementation, it is, perhaps, not surprising that they did not have capacity to engage with testing either.

Future research and publications

The IPE findings indicated that some pupils struggled to fully engage with the intervention (e.g., pupils who are shy or with certain SEN). Further research on ways to adapt the intervention for greater accessibility to all pupils may be beneficial.

Given the identified far transfer issue between FTM and reading attainment, it may be sensible to look at outcomes closer to the intervention. Based on the revised logic model, phonological awareness and understanding of numbers could be sensible outcomes to explore.

An additional overarching report on all five Learning About Culture interventions funded by EEF and the RSA will be published in 2021. This will include the three Key Stage 2 Learning About Culture interventions (Craft of Writing, Power of Pictures and the Young Journalist Academy) and the two Key Stage 1 Learning about Culture interventions (First Thing Music and Speech Bubbles). This report pools outcome data across the trials for a combined impact evaluation and synthesises IPE results across all interventions.

References

Bruning, R., Dempsey, M., Kauffman, D.F., McKim, C., Zumbrunn, S., 2013. Examining dimensions of self-efficacy for writing. *Journal of Educational Psychology* 105, 25.

Catterall, J.S., 2009. *Doing well and doing good by doing art: The effects of education in the visual and performing arts on the achievements and values of young adults*. Los Angeles/London: Imagination Group/I-Group Books.

Catterall, J.S., Dumais, S.A., & Hampden-Thompson, G., 2012. *The arts and achievement in at-risk youth: Findings from four longitudinal studies*. National Endowment for the Arts. [online] https://www.researchgate.net/publication/263479321_The_Arts_and_Achievement_in_At-Risk_Youth_Findings_from_Four_Longitudinal_Studies.

Courey, S.S., Balogh, E., Siker, J., & Paik, J., 2012. Academic music: Music instruction to engage third-grade students in learning basic fraction concepts, *Educational Studies in Mathematics*, 81(2), 251–278.

David, D., Wade-Woolley, L., Kirby, J.R., & Smithrim, K., 2007. Rhythm and reading development in school-age children: a longitudinal study. *Journal of Research in Reading*, 30(2), 169–183.

Duncan, D. J., 2007. *The relationship between creativity and the Kindermusik experience*. Ann Arbor, University of Central Missouri. 1448360: 61.

EEF, 2018. *Statistical analysis guidance for EEF evaluations*. Education Endowment Foundation, London.

Foregard, M., Schlaug, G., Norton, A., Rosam, C., Iyengar, U., & Winner, E., 2008. The relation between music and phonological processing in normal-reading children and children with dyslexia, *Music Perception* 25(4), 383–390.

François, C., Chobert, J., Besson, M., & Schön, D., 2013. Music training for the development of speech segmentation. *Cerebral Cortex* 23(9), 2038–2043.

Gerber A.S. & Green D.P., 2012. *Field experiments*. New York: W. W. Norton & Company.

Gordon, R.L., Fehd, H.M., & McCandliss, B.D., 2015. Does music training enhance literacy skills? A meta-analysis, *Frontiers in Psychology* 1 December. [online] https://www.frontiersin.org/articles/10.3389/fpsyg.2015.01777/full#B26

Gresham, F.M. & Elliott, S.N., 2011. Test review. *Journal of Psychoeducational Assessment* 29(3), 292–296.

Gromko, J.E., 2005. The effect of music instruction on phonemic awareness in beginning readers, *Journal of Research in Music Education* 53(3).

Hannes, K., 2011. Chapter 4: Critical appraisal of qualitative research. In J. Noyes, A. Booth, K. Hannes, A. Harden, J. Harris, S. Lewin, & C. Lockwood (eds.), *supplementary guidance for inclusion of qualitative research in Cochrane systematic reviews of interventions (version 1).* Retrieved from www.cqrmg.cochrane.org/supplemental-handbook-guidance

Henley, D., 2011. *Music Education in England*. Department for Education. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/175432/DFE-00011-2011.pdf

Kleiber, C., & Zeileis, A., 2008. *Applied econometrics with R*. Springer-Verlag, New York. ISBN 978-0-387-77316-2, https://CRAN.R-project.org/package=AER.

Neelands, J., Belfiore, E., Firth, C., Hart, N., Perrin, L., Brock, S., Holdaway, D., & Woddis, J., 2015. *Enriching Britain culture, creativity and growth: 2015 report by the Warwick Commission on the future of cultural value*. University of Warwick, Coventry.

Newman, M., Bird, K., Tripney, J., Kalra, N., Kwan, I., Bangpan, M., & Vigurs, C., 2010. *Understanding the impact of engagement in culture and sport: A systematic review of the learning impacts for young people*. CASE: The culture and sport evidence programme. [online] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/88447/CASE-systematic-review-July10.pdf

Nowell, L.S., Norris, J.M., White, D.E., & Moules, N. J., 2017. Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods* doi: 10.1177/1609406917733847

Patscheke, H., Degé, F., & Schwarzer, G., 2016. The effects of training in music and phonological skills on phonological awareness in 4- to 6-year-old children of immigrant, *Frontiers in Psychology* 7. [online] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5073325/

Ritchie, J., Lewis, J., McNoughton-Nichols, C., Ormston, R. (eds.), 2014. *Qualitative research practice*. 2nd edn. London: Sage

Sala, G. & Gobet, F., 29 July 2020. Cognitive and academic benefits of music training with children: A multilevel meta-analysis. *Memory and Cognition*. [online] https://rd.springer.com/article/10.3758/s13421-020-01060-2

See, B.H., & Ibbotson, L., 23 March 2018. A feasibility study of the impact of the Kodály-inspired music programme on the developmental outcomes of four to five-year olds in England. *International Journal of Educational Research* Volume 89, 2018, 10–21.

See, B.H., & Kokatsaki, D. 2016. *Impact of arts education on the cognitive and non-cognitive outcomes of school-aged children: A review of evidence*. [online] https://educationendowmentfoundation.org.uk/public/files/Presentations/Publications/Arts_Education_Review.pdf

Sterne, J.A.C., Savović, J., Page, M.J., Elbers, R.G., Blencowe, N.S., Boutron, I., Cates, C.J., Cheng, H-Y., Corbett, M.S., Eldridge, S.M., Hernán, M.A., Hopewell, S., Hróbjartsson, A., Junqueira, D.R., Jüni, P., Kirkham, J.J., Lasserson, T., Li, T., McAleenan, A., Reeves, B.C., Shepperd, S., Shrier, I., Stewart, L.A., Tilling, K., White, I.R., Whiting, P.F., & Higgins, J.P.T., 2019. RoB 2: a revised tool for assessing risk of bias in randomised trials. *The BMJ*, 366: l4898.

UK Government, 2019. *Get information about schools*. [online] Available through https://get-information-schools.service.gov.uk/ [Accessed in February 2019].

Appendix A: EEF cost rating

Figure A1: Cost rating

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: Progress in Reading Assessment (PIRA)

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%					
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%	3		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%					2
1	Design for comparison that does not consider selection on any relevant confounders	0.50 - 0.59	41-50%					
0	No comparator	>=0.6	>50%					

Threats to validity	Risk rating	Comments
Threat 1: Confounding	Low	Randomisation was appropriate, analysis code included in report and quality assured by independent researcher at BIT. Assessing imbalance at baseline limited due to data restrictions, namely no measure for prior attainment. For % FSM and EAL groups are balanced.
Threat 2: Concurrent Interventions	Low	Schools did offer music sessions as part of their BAU but there was no evidence that pupils took part in other extra music interventions that might influence the results of this trial.

Threat 3: Experimental effects	Low	Within school class randomisation which can sometimes lead to contamination, but this was discussed in the report and data collected with teachers in control condition indicated a low risk.
Threat 4: Implementation fidelity	High	Fidelity to the delivery model was low. Only 54% delivered sessions in the morning and (42%) delivered sessions 4-5 times per week. Compliance was also low. Only 40% of teachers met the minimum compliance threshold, which took into account delivery frequency and training attendance. CACE analysis indicated an effect size twice as large as that observed in the ITT analysis. This could mean the impact estimate is an underestimation.
Threat 5: Missing data	Moderate	Overall pupil-level attrition is high (28%) The rate is higher in the intervention group at 31% compared to 24% in the control. The proportion of data missing for various known reasons is comparable between groups, aside from refusing testing that was 6.1% in the treatment group and 4.6% in the control group. Additional analysis finds that FSM is predictive of missing PIRA scores, therefore it is unlikely that attrition is completely at random. FSM is controlled for in the model.The lack of NPD and pre-test data also causes some concerns here, though there is little more the authors could have done under the circumstances explained in the report.
Threat 6: Measurement of outcomes	Low	Assessors and markers blind to condition. PIRA is a standardised measure. However, observations of students with lower literacy skills struggled to engage with the test are noted but as this would apply to both conditions, this is unlikely to cause biased results.
Threat 7: Selective reporting	Low	The study differs from the original protocol, but the reasons are justified to a satisfactory extent. Selective reporting not apparent.

- **Initial padlock score:** 4 Padlocks – Randomised design with 0.19 MDES at randomisation and 28% attrition.
- **Reason for adjustment for threats to validity**: -1 – One high and one moderate risk, with the direction of likely biases unclear. No further adjustments required.
- **Final padlock score:** initial score adjusted for threats to validity = 2 Padlocks

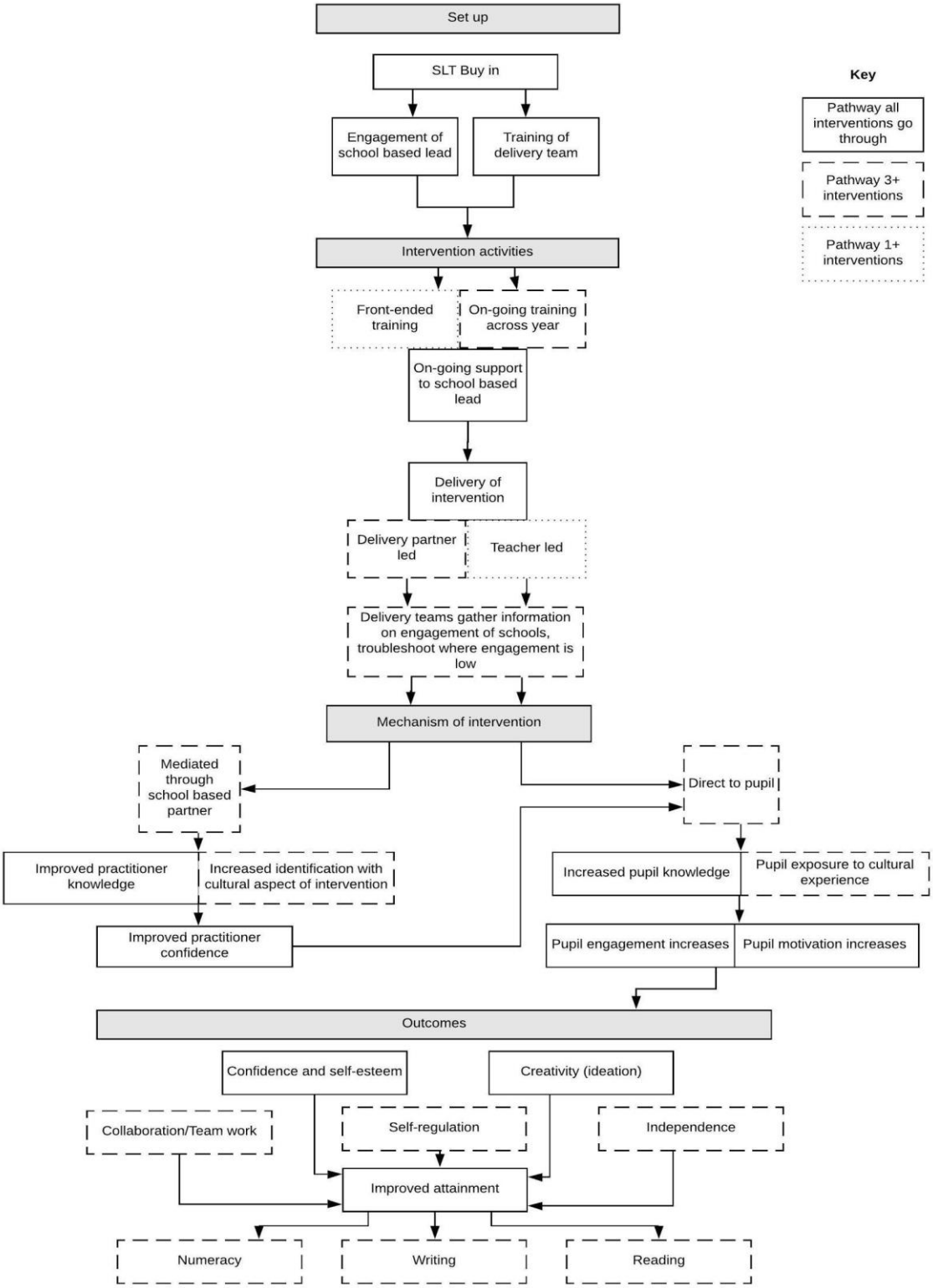
Appendix C: Effect size estimation

Table C1: Effect size estimation

			Intervention group		Control group		
Outcome	Unadjusted differences in means	Adjusted differences in means	n (missing)	Variance of outcome	n (missing)	Variance of outcome	Pooled variance
Raw PIRA Score	0.09	0.48	1116 (376)	6.80	1034 (478)	6.76	6.78
SSIS Score	−3.08	−2.20	433 (1059)	23.87	419 (1093)	26.55	25.22
WSEM3 Score	−0.02	−0.06	1025 (467)	1.64	906 (606)	1.63	1.64

Further appendices

Appendix D: Overarching flow diagram



Ethics & data protection

Appendix E: School information letter and MOU Agreement to participate in the evaluation of First Thing Music

First Thing Music – An Overview

A research project involving Teesside Year One classes and their teachers.

In partnership with Tees Valley Music Service, (TVMS), and the British Kodály Academy, (BKA), **First Thing Music** has been lucky enough to secure joint funding of £150,000 from the Education Endowment Foundation and the Royal Society of Arts.

https://www.thersa.org/action-and-research/rsa-projects/creative-learning-and-developmentfolder/learning-about-culture.

The drive behind the project is two-fold :

- to investigate the impact of daily Kodály-based musical activities on young children's academic and wider outcomes over the course of one full academic year
- to develop ways of giving non-specialist primary school teachers the confidence to adopt this kind of approach with music in their own classrooms.

We will be measuring both a primary and a secondary outcome through the trial; firstly looking at improvement in literacy; secondly we will be looking at self-regulation and disposition for learning.

Over the coming weeks, we will be selecting a team of Music Education Practitioners based in the area, all of whom will be trained by Lucinda Geoghegan and Zoe Greenhalgh of the BKA, to support and mentor 10 schools each.

Overall, we will be recruiting about 60 schools, who will be willing to take part in the trial in the following ways:

- by allowing their classes to be randomly allocated to either the intervention group or the control group;

- by supporting the allocated YR 1 teacher to participate in/lead all the 15 minute daily music sessions, and the CPD offered by the BKA team;
- and by their cooperation with the evaluators of the project – the Behavioural Insights Team at the Institute of Education, University College London.
- There will be a £200 fee to secure a place on the project, that will be put towards the cost of the printed resources.

TVMS, as partners in the First Thing Music team, will monitor and support this work, feeding in their knowledge of all the local schools across their 4 areas, (Redcar & Cleveland, Middlesbrough, Stockton, and Hartlepool), and other participating schools from the North East, hosting training sessions, and providing administrative back-up.

The preferred model is to work with schools that have a 2-class intake, (so that one class participates in the intervention, and the other class carries on as usual), though it may also be possible to work with some schools with a 1-class in-take, if those schools are happy to be randomised, and possibly be one of the schools that acts as the 'control' - ie just working as normal, without this particular intervention. (Post-project, the training and resources will be available for control schools that wish to take advantage of them.)

After all the recruiting, both of Practitioners and schools, (over Spring/Summer 2018), the trial will begin in earnest in September, for a full academic year. A resources pack will provide the guidelines for practitioners and teachers to follow closely – in order to keep the comparison between control and intervention groups as meaningful as possible. In July 2019 all the data will be collated, and by May 2020, the Education Endowment Foundation and the Royal Society of Arts plan to publish the results.

If your school would like to get involved in this project, then there is a window of opportunity between now and the end of April. We would be happy to set up a meeting or call you to present a fuller picture and answer any queries.

I do hope you can take advantage of this opportunity, and look forward to hearing from you very soon.

Lindsay Ibbotson

Project Lead

Memorandum of Understanding

Aims of the evaluation

The aim of this project is to evaluate the effect of *First Thing Music* on reading and social skills of year 1 pupils. *First Thing Music* is a structured music education programme based on the *Kodály* approach. Students will learn the basics of music through daily singing and musical games, their teachers being trained by members of the British Kodály Academy. The results of this project will make a contribution to understanding the value of music in improving pupil attainment.

The project

This evaluation is funded by the Education Endowment Foundation (EEF). *First Thing Music* will be delivered as a whole class intervention by class teachers trained by Kodály Music Specialists in association with the Tees Valley Music Service. Lindsay Ibbotson will be leading this team.In this evaluation, the *First Thing Music* programme will run over the course of the school year. The project will be delivered in the participating schools to reception pupils who will be moving into year 1 for the 2018/19 school year. The impact of three terms of *First Thing Music* will be evaluated using a randomised controlled trial (RCT), by randomly assigning classes within schools to a 'control' group who will not participate in the *First Thing Music* programme, and a 'treatment' group who will.

During this project, you will be contacted by both the *First Thing Music* team, (including Tees Valley Music Service and Kodály Music Specialists), hereafter referred to as the 'Project Team', who are responsible for programme delivery, support and resources; and by the Behavioural Insights Team, the Institute of Education UCL and the Royal Society of Arts (RSA), hereafter referred to as the 'Evaluation Team', who are carrying out an independent evaluation of the project.

Structure of the evaluation

Year 1 classes will be randomly assigned by the Evaluation Team to either receive the *First Thing Music* programme or not. Pupils in the treatment classes selected to receive the programme will take part in daily 15-minute music sessions throughout the course of the school year. Children in the classes which are allocated to the control group will continue as normal. Random allocation of classes to the control or treatment arms is essential to the evaluation. It is important that schools understand that classes will be assigned to either group and that they are agreeing to abide by this assignment.

Academic attainment with respect to reading will be the primary outcome measure in this impact evaluation, and will be measured using the Progress in Reading Assessment (PIRA) by *Rising Stars*. PIRA assessment will be collected by Research Assistants (RAs) employed by the Evaluation Team after the programme concludes in summer 2019. A survey measuring child social skill will also be completed by teachers at this time.

Use of Data

All data, including pupils' test responses and any other pupil data (e.g. Free School Meal eligibility and EYFSP scores), will be treated with the strictest confidentiality. Pupil assessments will be marked by RAs hired by the Evaluation team, or the test publisher. Pupil data will be matched with the National Pupil Database and shared with the Institute of Education, the Department for Education, the EEF, FFT Education and in an anonymised form the UK Data Archive. All results will be anonymised so that no schools or individual pupils will be identifiable in the report or dissemination of any results. Confidentiality will be maintained and no one outside the Project Team and Evaluation Teams will have access to the database.

Responsibilities

The First Thing Music team agrees:

- To devise, implement and support music sessions as agreed, through the Kodály music specialists;

- To plan and implement training and evaluation sessions, through the Kodály music specialists, for teachers delivering the programme;

- To ensure Kodály music specialists have a current DBS enhanced disclosure, and the relevant safeguarding training certificates;

- To report any child protection concerns to the school's designated person;

Training notes

The Teacher training is scheduled for Tuesday 11th September 2018, 9.30 – 3.30 p.m.

In addition there will be 6 half-termly CPD sessions with the Music Practitioners and the participating teachers from the assigned schools on the following dates:

- 3/10/18 1.30 – 4.30 p.m.;
- 21/11/18 1.30 – 4.30 p.m.;
- 16/1/19 1.30 - 4.30 p.m;
- 27/2/19 1.30 – 4.30 p.m;
- 8/5/19 1.30 - 4.30 p.m;
- 26/6/19 1.30 – 4.30 p.m;
- There will also be a celebratory event on 27/6/19 (time to be confirmed)

The Music Practitioner will plan and deliver 15 minute music sessions, based upon the contents of the First Thing Music pack, during two full weeks in September, (with a seven day gap between the two delivery weeks during which the class teacher will be encouraged to try leading some sessions for themselves).

Thereafter, the Music Practitioner will act as a 'musical mentor' for the class teacher, providing weekly visits as required throughout the academic year, along with guidance and any other necessary support.

The Behavioural Insights Team, Institute of Education UCL and The Royal Society of Arts agree:

- To conduct the random allocation of classes to trial arms;
- To collect class and pupil level data; including UPNs, first and last names, DoB, FSM status, classroom teacher, EAL status and gender for all children;
- To collect assessments of aforementioned outcome measures;
- To store all data safely and securely;
- To analyse data from the project in order to produce impact estimates;
- To conduct the process evaluation, including analysis and reporting from this;

To produce end of project evaluation report to be published by the Education Endowment Foundation.

Primary School agrees:

- To enable the *First Thing Music* sessions to take place in a space suitable for practical group music sessions within the times discussed and agreed;

- That 2018/19 Year 1 teachers assigned to the treatment group will attend all *First Thing Music* training sessions (and where necessary for the six afternoon sessions are given cover to attend) and classroom sessions, and to

create a suitable contingency if an allocated staff member cannot attend due to sickness (etc.). Acceptable contingencies include providing an alternative member of staff or rescheduling of session by agreement with *First Thing Music*;

- To allocate time for the Head Teacher to participate in and make observations on at least one session per half-term;
- To have gained agreement from 2018/19 Year 1 teachers and teaching assistants for collaborating with the Evaluation team when required, chiefly with respect to completing surveys concerning the social skills of children in the study, and allowing a reading measure to be collected during class time;
- To make Year 1 and senior staff available for the purposes of interviews in order to complete a qualitative assessment of the programme, and to complete a brief survey if requested;
- To distribute information forms to parents of reception year children in the 2017/18 school year, and to keep a record of any parents who exercise their right to withdraw from their child's data being used as part of the trial;
- For those children whose parents do not choose to withdraw, to provide the Evaluation team with the necessary pupil data two weeks after the letter is sent out.
- To abide by the random assignment of classes to the *First Thing Music* programme, and NOT attempt to use the programme in the class allocated to the control group;
- To nominate a staff member to act as a primary contact, which involves attending planning and progress meetings in order to ensure the success of the project taking responsibility for the timely distribution of consent forms to parents, provision of pupil data records, and scheduling of any Evaluation team visits for the purpose of conducting assessments or performing qualitative research.

Contact Details

Thank you for agreeing to take part in this research. Please do not hesitate to get in touch if you have any questions about the *First Thing Music* programme or the evaluation itself.

[CONTACT DETAILS REDACTED]

Agreement

I agree for my school to take part in the evaluation of *First Thing Music* research project and I accept the requirements outlined in this MoU.

Appendix F: Parent information letter, withdrawal form, and data privacy notice

Dear Parent/Carer,

Your child's school is taking part in a research project that aims to improve social skills and reading ability. The *First Thing Music* programme, in association with Tees Valley Music Services, uses Kodály-based musical activities to promote children's communication, confidence and wellbeing. The Kodály approach has a long track record of helping children's musical development and seems to be beneficial for their confidence. The approach introduces children to musical concepts through experiences such as listening, singing, and movement. The programme will run in 2018/19 school year.

As this is a research project, not all children in the school will actually end up taking part in First Thing Music, but will carry on with their normal activities. Whether your child's class will be part of the programme will be decided randomly by evaluators (The Behavioural Insights Team, UCL Institute of Education and The Royal Society of Arts) in order to look at how effective *First Thing Music* has been.

As part of measuring the success of *First Thing Music*, children will complete a short classroom-based reading assessment in summer 2019. Staff hired and trained by the evaluators will conduct these assessments. Also in summer 2019, a survey concerning your child's social skills will also be completed by their teacher. These assessments will not be used to monitor teacher or school performance, and are only collected to help us understand how much *First Thing Music* benefits children.

Your child's name and other data held by the school, alongside their early years foundation stage profile (EYFSP) and reading assessment scores, will be collected by the evaluators in conjunction with Tees Valley Music Service and contracted test markers. No information that can identify individual children will be made available to anyone outside these teams and your child's school. This data will be kept under password protection in an offline format. We will not use your child's name or the name of the school in any report arising from the research, and no information that could otherwise identify your child will be made public.

We will also obtain your child's UPN (Unique Pupil Number) to allow longer term understanding of whether this programme worked. We will then save this information in a data format that will prevent anyone from identifying your child. This data will then by linked with the National Pupil Database (held by the Department for Education) and shared with the project team, the Department for Education, Education Endowment Foundation (EEF), EEF's data contractor FFT Education and in an anonymised form to the UK Data Archive.

This research has been reviewed and approved by the ethics committees of UCL Institute of Education.

If you have any questions you would like to ask, please contact [REDACTED]

If you are happy for information about your child to be used in the *First Thing Music* research project you do not need to do anything. Thank you for your help with this research, your support is much appreciated. You and your child can withdraw from the study at any time by emailing the evaluators.

If you **DO NOT** want information about your child to be used to understand whether the *First Thing Music* programme can help children to improve their reading and social skills, please complete the enclosed form and return it to your child's school by [date]. If you do this then no information about your child will be shared with the evaluation or project teams at any point during the project. This will **not** affect your child's ability to take part in any *First Thing Music* activities that happen in their school.

With thanks from the Research and Evaluation Teams

If you are happy for your child to participate in the research on whether this programme improves reading and social skills, you **DO NOT** need to return this form.

I **DO NOT** wish data about my child to be collected as part of this research.

Child's name: ……Date of birth: ……

Child's class Teacher: …….

School:……

Parent name (BLOCK CAPITALS) ……

Parent signature: ……..

Date ……

(Please return the completed form to your child's class teacher.)

EEF/RSA EVALUATION OF FIRST THING MUSIC

Data Privacy Notice

Dear Parent,

We've previously been in touch because your school is taking part in a project funded by the Education Endowment Foundation (EEF) to understand the potential benefits of First Thing Music. As part of that information, you were given the opportunity to tell your school not to pass any data about your child to us to be used as part of this project. **Please rest assured that if you contacted the school to make this request, nothing in this letter changes that.**

We wanted to get in touch again to provide you with further details about the way will be handling pupils' data as part of this project. **It is very important to us that that we do this responsibly and providing these details are an important part of that.** They are also important in fulfilling our responsibilities under the UK's data protection laws, which we take very seriously. These require us to provide you with some specific information about **our plans** and **your rights**.

Some of this description involves rather technical terms, which we've left in so you know the official concepts we are talking about. We've tried to keep the explanations as simple as possible. If we haven't managed that well enough and you have any questions now, or at any point during this project, then you should contact a member of the team with the first point of call being Louise Jones at BIT [email redacted] and Jake Anders at UCL [email redacted]

Our Plans

- Using pupils' data as part of research is not something we do without thinking about it. Under data protection law, we require a "lawful basis" for the data processing that we carry out. UCL will be using the lawful basis known as the "public task" basis, while the lawful basis BIT are using is known as the "legitimate interests" basis (it is different at UCL and BIT because UCL is a university). To use the "legitimate interests" basis, we must consider why this is a legitimate interest and inform you of this. Here, it is because our work is for the purpose of promoting the education or well-being of children in England and couldn't be achieved without analysing these test scores. We balance that against your and child's rights by providing you with the right to object to our use of your child's data in this way.
- Your child's name and other data held by the school, alongside the test scores we will collect, will be collected and processed by us for the purpose of understanding how participating in this project has affected their learning. No information that can identify individual children will be made available to anyone outside these teams and your child's school (with an important exception, as explained below). We will also obtain your child's UPN (Unique Pupil Number) from the school to allow us to link up our data with the National Pupil Database (held by the Department for Education) and other official records to understand whether being part of this project is linked with test scores when they are older. This involves us sharing data with the Department for Education (part of the UK Government), the Education Endowment Foundation (who funded the trial), EEF's data processor FFT Education and (in a form that will prevent anyone from identifying your child) to the UK Data Archive.
- The data we hold will be kept securely at all times, transferred using secure (encrypted) methods, and kept on secure computer systems at UCL and BIT's offices under password protection. We will never use your child's name or the name of the school in any report arising

from the research, and no information that could otherwise identify your child will be made public.

- Pupils' personal data will be processed by us only for the purposes of this research project. Once that is complete then the data will be securely destroyed from our computer systems. Personal data will certainly not be more than 10 years, in line with UCL's policy on storing research data.

Your rights

- Under data protection law, you have a right to be informed about our plans. This letter, as well as the information that you previously received from your school, are all part of this. You also need to know exactly who is involved in the data processing. In legal terms, University College London (UCL) and the Behavioural Insights Team (BIT) are considered joint data controllers for this project. The law requires our organisations to have named Data Protection Officers, who are ultimately responsible for overseeing data processing that goes on in their respective organisations. UCL's Data Protection Officer is Lee Shailer, who can be contacted via **data-protection@ucl.ac.uk**. BIT's Data Protection Officer is Lizetta Lyster, who can be contacted via **dpo@bi.team**. You should contact these individuals if you have any complaints about how we are processing data. However, if you remain unsatisfied, you may wish to contact the Information Commissioner's Office (ICO). Contact details, and details of data subject rights, are available on the ICO website at: **https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/individuals-rights/**
- As noted above, we provided you with the right to object to data processing before schools handed any information over to us. You can also contact us at any point during the project to request information we hold about your child, to request rectification of any information that is incorrect, to stop using their data as part of the project or to destroy their data. If you wish to make such a request or ask any questions about it then please contact us. The best place to start is to contact Louise Jones at BIT [email redacted] and Jake Anders at UCL [email redacted].

Once again, we are extremely grateful to you for supporting this project. We hope to learn a lot about the role cultural and arts education can play in supporting pupils' learning.

Appendix G: Adapted Writing Self-Efficacy Measure (WSEM3)

My ideas

Name: _____

Instructions: Put a tick in one box for each statement below.

	I don't do this much	I do this sometimes	I do this lots
1. I think of new ideas.			
2. I tell my ideas to other people.			
3. I do things with my ideas.			

Thank you for helping us by thinking about your ideas.

Appendix H: Randomisation syntax

```
global dta "[FILE PATH]"
global raw "[FILE PATH]"

use "${dta}data for randomisation", clear

*randomise
//randomising at class level and merging with pupil level data

 preserve
        set seed 54531 //set seed
        keep urn class_id numforms
        duplicates drop

        bysort urn (class_id) : gen rand = runiform()
        bysort urn (rand) : gen n=_n
        bysort urn (rand) : egen treat1 = seq() if n<3 & numforms!=1, f(0) t(1) //randomise two forms per school
                tab urn treat1 //check
        sort treat1 rand
        egen treat2 = seq() if missing(treat1), f(0) t(1) //randomise one form schools and third form in three form schools
                tab urn treat2 //check

        egen treat = rowtotal(treat*) //combine the two
                tab treat //check
                drop treat1 treat2 n rand

        tempfile temp
        save "`temp'", replace
 restore
```

```
 merge m:1 class_id using "`temp'", keepusing(treat) //merge with pupil data

            drop _merge


/*On 30/07/2018 we found that two schools which were supposed to be randomised into the

trial were not due to duplicate URNs and the files being overwritten (multiple schools from same MAT and shared a URN
in the files). We randomise them below.*/


clear

 cd "${raw}" // spreadsheets of the overwritten schools that were not included in randomisation

 fs *.xls*


 *[cleaning lines omitted -- same cleaning as that which preceded the 'data for randomisation' file]


*checks

 tab numforms, m // one school is one form; one school is three form


/*we currently have 8 one form schools in control and 10 in treatment - one form school should

be assigned to control; In three form schools we have 18 classes in control and 15 in treatment -

two out of three classes in the three form school should be assigned to treatment*/


 set seed 54531

 bysort class_id : gen rand = runiform() if numforms==3 & _n==1


 sort rand

 egen treat = seq() if !missing(rand), f(1) t(0)

 replace treat = 0 if numforms == 1
```

Appendix I: Analysis syntax

Primary analysis

```
model_P1 <- lm(pira_score ~ treat + fsm + eal + eal_class_avrg + fsm_class_avrg + anonschoolid, data = DT)

coeftest(model, vcov=vcovHC(model, cluster="anonclassid"))
```

Secondary analysis

```
model_S1 <- lm(ssis_score ~ treat + fsm + eal + eal_class_avrg + fsm_class_avrg + anonschoolid, data = DT)

coeftest(model_S1, vcov=vcovHC(model_S1, cluster="anonclassid"))

model_S2 <- lm(wsem3_score ~ treat + fsm + eal + eal_class_avrg + fsm_class_avrg + anonschoolid, data = DT)

coeftest(model_S2, vcov=vcovHC(model_S2, cluster="anonclassid"))
```

CACE model

```
CACE_model <- ivreg(pira_score ~ used_ftm + fsm + eal + anonschoolid + eal_class_avrg + fsm_class_avrg | treat +fsm + eal+ anonschoolid + eal_class_avrg + fsm_class_avrg,

        data = DT[is.na(pira_score)==FALSE])

c <- coeftest(CACE_model, vcov = vcovHC(CACE_model, cluster="anonclassid"))
```

Subgroup analysis – FSM

PIRA

```
fe_model <- lm(pira_score ~ treat*fsm +eal_class_avrg + fsm_class_avrg+ anonschoolid, data = DT[is.na(pira_score)==FALSE])

c <- coeftest(fe_model, vcov=vcovHC(fe_model, cluster="anonclassid"))
```

SSIS

```
fe_model <- lm(ssis_score ~ treat*fsm +eal_class_avrg + fsm_class_avrg+ anonschoolid, data = DT[is.na(ssis_score)==FALSE])

c <- coeftest(fe_model, vcov=vcovHC(fe_model, cluster="anonclassid"))
```

WSEM3

```
fe_model <- lm(wsem3_score ~ treat*fsm +eal_class_avrg + fsm_class_avrg+ anonschoolid, data = DT[is.na(wsem3_score)==FALSE])

c <- coeftest(fe_model, vcov=vcovHC(fe_model, cluster="anonclassid")) #no sign interaction
```

Appendix J1: Teacher interview guides

First Thing Music: Interviews with Classroom Teacher

The interviews should last around 30 minutes. The timings given for each section are a guide - you may spend longer or shorter on each section. Lead questions are presented in bold, with potential follow-up questions presented in a non-bold typeface. As the interviews are semi-structured, not all questions need to be asked and they do not need to be asked in order. The interviewer should be responsive to what the interviewee, following the direction of the conversation and following-up with additional questions as needed.

Main objective	Purpose of section	Guide timings
1. Introduction	Explains the purpose and 'ground rules' of the interview.	3 mins
2. Background context	Allows the participant an opportunity to settle into the interview, as well as providing some background to the school and the context in which the First Thing Music programme is being delivered.	5 mins
3. Delivery experience	This section will focus on understanding the perceived quality of the intervention, as well as experiences of the programme's delivery, including barriers and facilitators to delivery.	5 mins
4. Pupil engagement and programme mechanisms	To explore the teacher's perception of pupils' engagement in the sessions and the positive and negative impact of the programme, together with the mechanisms that brought about any impact identified.	5 mins
5. Training and support	To understand teacher's experience of the training they received to deliver the programme, and the support they have received to deliver the programme, including from the school's SLT.	5 mins
6. Close	Thank you and close.	2 mins
Observation of FTM session		
7. Reflection following observation of FTM session	Where possible, this will be an opportunity to follow-up on any areas of interest arising from the observation.	5 mins

1. Introduction	3 mins

Introduction: ● Introduce yourself ● Introduce BIT and IOE – explain that we are independently evaluating the First Thing Music programme, which is one of five programmes that are part of the Cultural Learning programme that is jointly funded by the Education Endowment Foundation and Royal Society of Arts. **Aims of this interview:** We are here to learn more about how the First Thing Music programme has worked in your class. We're interested in what involvement you have had with the programme, what has helped the programme to work, and what the challenges have been. We'd also like to understand any impact the programme has had on your school, particularly pupils in your class. **This interview:** ● Should take no more than 30 minutes ● Stress that you want to understand the intervention from their point of view. No answers are right or wrong – and we are not here to judge the decisions made or views held by the interviewee. **Anonymity and privacy:** ● All information gathered will be in strict confidence, unless there are concerns about safeguarding. When we write up the research we will ensure that no one is identifiable from any reporting. ● Explain that if at any point they feel uncomfortable or prefer not to answer a specific question they can just say so. ● Explain that it is their choice whether they take part in the interview and they can end the interview at any point, without giving a reason. **Recording:** 1. Explain that recording enables us to have an accurate record of what was said, which can be typed up for analysis alongside other interviews. We may also use quotes from this interview, but these will be included in a way that means no individual or school is identifiable. 2. Check if they have any questions about the interview. If they are happy to go ahead, obtain verbal permission to digitally record and take notes (written permission should already have been obtained). 3. Once you have consent, start the voice recorder. 4. State interview number/participant ID	Orientates respondent and gets them prepared to take part in the discussion. Outlines the 'rules' of the interview.
2. Background context	5 mins

How long have you been teaching? **How long have you been working at this school?** ● What would you say are the school's main strengths and challenges? **What musical experience do you have?** **What are your views about music teaching in primary schools?** **How did you become involved with the First Thing Music project?** **What were your thoughts about the First Thing Music programme when you first heard of it?**	Allows the participant an opportunity to settle into the interview, as well as providing some background to the school and the context in which the First Thing Music programme is being delivered.
3. Delivery experience	5 mins
When did you start your First Thing Music sessions? **How have you found the First Thing Music sessions?** ● What do you like about the First Thing Music programme? ● What have the main challenges been? ● What do you think could have helped to overcome these? ● What recommendations do you have about how the programme could be improved? **What musical lessons or activities do pupils in Year One usually do?** ● How would you say these compare to the First Thing Music approach? **How have the First Thing Music sessions been organised and timetabled in your curriculum?** ● When in the day do you typically run the sessions? Why is that? ● How long do they typically last? What do you think about that length of session? ● Where do the sessions typically take place? How have you found using this space for the sessions? **Can you tell me about how you've found working with the music specialist?** ● How have they been with the pupils? ● To what extent have you been able to share your knowledge with one another?	This section will focus on understanding the perceived quality of the intervention, as well as experiences of the programme's delivery, including barriers and facilitators to delivery.

4. Pupil Engagement and Programme Mechanisms	5 mins
How have the children in your class responded to the First Thing Music sessions? ● Could you describe the pupils who have got the most out of the sessions? Why do you think this is? ● Could you describe any children who got less out of the sessions? Why do you think this is? **What changes, if any, have you seen in your pupils as result of taking part in the First Thing Music programme?** ● If not mentioned, probe for effect on confidence, musical skills, literacy, creativity and social skills, e.g. Can you describe any effect of the programme on pupils' creativity? ● What is it about programme that you think has helped them to develop their (e.g. confidence)? ● Can you describe any negative consequences of the programme for pupils? **What effect do you think taking part in the First Thing Music programme has had for you?** ● What was it about the programme that you think led to these changes?	To explore the the teacher's perception of pupils' engagement in the sessions and the positive and negative impact of the programme, together with the mechanisms that brought about any impact identified..
5. Training and support	5 mins
What were your views about the CPD sessions that you attended as part of the First Thing Music programme? ● What did you cover in the CPD sessions? ● What worked well about the CPD sessions? ● What did you find challenging about the CPD sessions? ● To what extent did the CPD sessions support you to deliver the First Thing Music programme? ● If you could change anything about the training, what would it be? **To what extent have the SLT been involved in the First Thing Music programme?** ● Can you describe any support you've received from SLT to deliver the programme? ● Can you describe any resources they've made available to deliver the programme?	To understand teacher's experience of the training they received to deliver the programme, and the support they have received to deliver the programme, including from the school's SLT.

6. Close	2 mins
Overall, would you recommend First Thing Music to other schools? ● Why/Why not? **Was there anything else that you were hoping to discuss that we haven't yet had a chance to talk about?** Thank the interviewee for their time and reassure them of the confidentiality of their responses, as explained at the beginning of the interview.	Thank you and close.

7. Reflection following observation of FTM session	5 mins
Below are some exemplar questions, however, the interviewer may want to add some additional questions based on their observations. Probe for their perception of student engagement; implementer support; and mechanisms. **What are your reflections on how that session went?** ● What do you think went well? ● What do you think the challenges were? ● Which students engaged particularly well? How/why? ● Which students engaged less well? How/why? ● If you could go back and do that session again, what might you do differently? **I noticed this [describe significant moment in the lesson]: what was your view about what was happening?** **How does the observed session compare to other First Thing Music sessions that you and your class have been involved with?** Probe specific areas of interest from your observation/ in relation to the review questions/ thinking about how this school is running First Thing Music sessions in comparison to other schools. **Could you describe any ways that the sessions have changed over time?** **What are your plans for future First Thing Music sessions?**	Opportunity to follow-up on any areas of interest arising from the observation.

Appendix J2. Senior leadership team interview guide

First Thing Music: Interviews with Senior Leadership Team

The interviews should last around 20 minutes. The timings given for each section are a guide - you may spend longer or shorter on each section. Lead questions are presented in bold, with potential follow-up questions presented in a non-bold typeface. As the interviews are semi-structured, not all questions need to be asked and they do not need to be asked in order. The interviewer should be responsive to what the interviewee, following the direction of the conversation and following-up with additional questions as needed.

Main objective	Purpose of section	Guide timings
1. Introduction	Explains the purpose and 'ground rules' of the interview.	3 mins
2. Background context	Allows the participant an opportunity to settle into the interview, as well as providing some background to the school, so that we understand more about the context in which the First Thing Music programme is being delivered.	2 mins
3. Engagement	To understand the extent to which the school, and particularly the SLT, has engaged with and supported the programme.	4 mins
4. Delivery experience	This section will focus on understanding the perceived quality of the intervention, as well as the experience of the programme's delivery, including barriers and facilitators to delivery. We will also try and disentangle what is 'business as normal' within the school in terms of music, and whether this has changed as a result of the First Thing Music programme.	5 mins
5. Mechanisms of change	To explore the SLT member's perception of the positive and negative impact of the programme, particularly for pupils, and the mechanisms that brought about any change identified.	4 mins
6. Close	Thank you and close	2 mins

1. Introduction	3 mins
Introduction: ● Introduce yourself	Orientates respondent and gets them prepared to take part in the discussion.

● Introduce BIT and IOE – explain that we are independently evaluating the First Thing Music programme, which is one of five programmes that are part of the Cultural Learning programme that is jointly funded by the Education Endowment Foundation and Royal Society of Arts. **Aims of this interview:** We are here to learn more about how the First Thing Music programme has worked in your school. We're interested in what involvement you have had with the programme, what has helped the programme to work and what the challenges have been. We'd also like to understand any impact the programme has had on your school, particularly pupils in (the) First Thing Music class(es). **This interview:** ● Should take no more than 20 minutes ● Stress that you want to understand the intervention from their point of view. No answers are right or wrong – and we are not here to judge the decisions made or views held by the interviewee. **Anonymity and privacy:** ● All information gathered will be in strict confidence, unless there are concerns about safeguarding. When we write up the research we will ensure that no one is identifiable from any reporting. ● Explain that if at any point they feel uncomfortable or prefer not to answer a specific question they can just say so. ● Explain that it is their choice whether they take part in the interview and they can end the interview at any point, without giving a reason. **Recording:** 1. Explain that recording enables us to have an accurate record of what was said, which can be typed up for analysis alongside other interviews. We may also use quotes from this interview, but these will be included in a way that means no individual or school is identifiable. 2. Check if they have any questions about the interview. If they are happy to go ahead, obtain verbal permission to digitally record and take notes (written permission should already have been obtained). 3. Once you have consent, start the voice recorder. 4. State interview number/participant ID	Outlines the 'rules' of the interview.
2. Background context	2 mins

How many years have you been teaching? **How long have you been working at the school?** **Could you tell me about your role at the school?** **Could you tell me about the school?** ● What would you say are the schools' main strengths? ● What would you say are some of the school's biggest challenges? **What are you views about music teaching in primary schools?**	Allows the participant to settle into the interview, as well as providing some background to the school, so that we understand more about the context in which the First Thing Music programme is being delivered.
3. Engagement	4 mins
How did your school become involved with the First Thing Music programme? ● When did you first hear about the First Thing Music programme? ● Why did your school decide to get involved in the programme? ● What initial expectations did you have for the programme? **Can you describe what your involvement with the First Thing Music programme has been?** ● Can you describe any support that you or other members of SLT have provided to the programme? ● Can you describe any resources that you've made available to enable the programme to run? **Have you sat in on any First Thing Music sessions?** If yes: ● Can you describe any sessions you've sat in on? ● What did you think of the session(s)? [probe for strengths and suggested areas for improvement] If no: ● Were there any particular reasons that you didn't sit in on a session? (especially if headteacher as sitting on one session per half term is in MoU) ● What do you understand about what the class have been doing in their First Thing Music sessions?	To understand the extent to which the school, and particularly the SLT, has engaged with and supported the programme

4. Delivery experience	5 mins
Can you tell me what musical lessons and activities are typically on offer for pupils at your school? ● What musical lessons and other activities would pupils in Year 1 typically receive? ● Have you continued with these activities? ● What do you think about the First Thing Music programme in comparison to these other activities? ● Can you tell me any recommendations you have changes to the programme? **Thinking specifically about the delivery of the programme, how have you found having the First Thing Music programme running in your school?** ● What has worked well about the delivery of the programme? ● What have been the main challenges to delivering the programme? ● What have you done to try and overcome these challenges? ● How has the school found accommodating the daily First Thing Music sessions into the school day? ● Knowing what you know now, would you approach anything about the programme delivery differently?	This section will focus on understanding the perceived quality of the intervention, as well as the experience of the programme's delivery, including barriers and facilitators to delivery. We will also try and disentangle what is 'business as normal' within the school in terms of music, and whether this has changed as a result of the First Thing Music programme.
5. Mechanisms of change	4 mins
How have pupils in [intervention class] found the First Thing Music sessions? ● To what extent have pupils engaged with the sessions? ● What effect do you think taking part in the First Thing Music programme has had for pupils? ● What was it about the programme that you think led to [change described]? ● Can you describe any pupils that you think the programme is particularly helpful for? ● Can you describe any pupils that you think the programme is less suited to? **What effect, if any, do you think that the programme has had for the teacher of [intervention class]?** ● Can you think of any other impacts that the programme has had for the [intervention class] teacher? **Can you describe any effect of the programme more broadly within the school?**	To explore the SLT member's perception of the positive and negative impact of the programme, particularly for pupils, and the mechanisms that brought about any change identified.

• Can you describe any effect of the programme on other pupils within the school? • Can you describe any effect of the programme on other teachers within the school? • Has [intervention class] doing First Thing Music sessions affected music provision within the school more generally? If yes, how?	
6. Close	2 mins
Overall, would you recommend First Thing Music to other schools? • Why/Why not? **Was there anything else that you were hoping to discuss that we haven't yet had a chance to talk about?** Thank the interviewee for their time and reassure them of the confidentiality of their responses, as explained at the beginning of the interview.	Thank you and close

Appendix J3. Guide for informal discussions with pupils

First Thing Music: Discussions with Pupils

Short discussions, lasting around five minutes will be conducted with pupils happy to speak to a researcher. These will happen during lesson time in the classroom setting. The researcher(s) will take fieldnotes and therefore the conversations will not be audio recorded.

Aims of discussions with pupils

To better understand pupils' responses to and engagement with First Thing Music.

Before conducting discussions

- The researcher should find out from the teacher in advance which pupils' parents have consented for them to take part in the research, and whether there are pupils who have particular communication needs, where the researchers will need to adapt their approach.
- The teacher should explain to the class: who the researcher is and what they are going to be doing; the purpose of the research; and reassure pupils that they do not have to talk to the researcher if they would prefer not to.
- Before speaking to a pupil, ask whether they are okay to speak to you and ensure that their body language indicates that they are happy to talk to you.

Discussion guide questions

Did you enjoy your First Thing Music lesson today?

- What did you like about it?
- What did you dislike about it?
- Is there anything you would like to do differently/do more of/do less of in your First Thing Music lessons?

Can you tell me about some of the other songs you've learnt in your First Thing Music sessions?

- Which one is your favourite?

What do you think of the music teacher who comes in to help with the First Thing Music lessons sometimes?

What music did you do before your First Thing Music lessons?

- What do you prefer?
- Why is that?
- Do you do any music outside of school? If yes:
    - What music do you do outside of school?
    - Do you like First Thing Music or [whatever music they've described doing] more?
    - Why is that?

Is there anything else you want to tell me?

Thank you for talking to me today.

Appendix K: Observation pro forma

First Thing Music: Observation Proforma

This proforma may be completed during the observation, or more extensive notes may be made during the observation, i.e. in a notepad, and then summarised on this form.

Date and time: Name of school and class: Number of pupils present: Name of observer: Other adults present: Notes	
Description of setting, space and resources used for the session, e.g. type of room; types of musical instruments available; pupils sitting on chairs at tables or other arrangements.	
What is covered in the introduction to the session, e.g. are the aims and expectations covered?	
What activities are carried out as part of the session, e.g. nature of small group activities and/or whole class activities.	
What is the balance of leadership of the classroom music activities between the teacher and the music specialist?	
How confident is the teacher in delivering the musical aspects of the session, e.g. did they sing along with the children or remain observing on the sidelines?	
To what extent are children engaged in the session, e.g. are they focused on listening and performing, or are they distracted and talking to their peers about unrelated things?	
What is the behaviour of the pupils like, e.g. mainly on-task; mainly disruptive; individual children not on-task	

but the majority engaged?	
To what extent do children seem able to understand and follow the session? What difficulties do children who struggle seem to have? How are they supported?	
What strategies are used to account for the needs and abilities of different pupils?	
Other	

Appendix L. Raw treatment condition survey data

Please note: some percentages may add up to slightly above or below 100% due to rounding

1.1 What is your role at the school?

Response	Number (%)
Class teacher	47 (98%)
Other members of SLT (e.g. Head of Key Stage 2)	2 (4%)
Other (please specify):	
- SENCO/SENDCO	3 (6%)
- Class teacher and HLTA	1 (2%)
- English Lead Key Stage 1	1 (2%)
- Music Lead	1 (2%)

N.B. Total percentage is greater than 100% because respondents could select more than one role.

2.1 How important do you think it is to include music in the primary school curriculum?

Response	Number (%)
Very important	33 (69%)
Somewhat important	13 (27%)
Neither important nor unimportant	1 (2%)
Somewhat unimportant	1 (2%)

2.2 I have a natural aptitude for music

Response	Number (%)
Disagree	17 (35%)
Agree	13 (27%)
Neither agree nor disagree	10 (21%)
Strongly agree	6 (13%)
Strongly disagree	2 (4%)

2.3 I am confident in performing music to my pupils

Response	Number (%)
Agree	23 (48%)

Strongly agree	9 (19%)
Neither agree nor disagree	8 (17%)
Disagree	6 (13%)
Strongly disagree	2 (4%)

2.4 I am confident in teaching music to my pupils

Agree	27 (56%)
Neither agree nor disagree	9 (19%)
Disagree	6 (13%)
Strongly agree	6 (13%)

2.5 I understand core music principles

Response	**Number (%)**
Agree	22 (46%)
Neither agree nor disagree	12 (25%)
Strongly agree	7 (15%)
Disagree	5 (10%)

Strongly disagree	2 (4%)

3.1 Did you attend training as part of the First Thing Music programme?

Response	Number (%)
Yes, I attended all sessions available to me	33 (69%)
Yes, I attended some of the sessions available to me	15 (31%)

3.2 Reasons for not attending training

Response	Number (%)
I had competing school commitments	5 (10%)
I could not find cover	5 (10%)
I could not travel the distance required	4 (8%)
I had personal commitments	2 (4%)
I did not receive time off in lieu	1 (2%)
Cost of travel and/or subsistence	1 (2%)
I was unwell	1 (2%)

SLT were not supportive of me going	1 (2%)
Other (please specify):	
- Would have not got back to school in time for parents' evening. Alternative venues were too far away.	1 (2%)
- Not allowed to join until January	1 (2)%
- Maternity leave	1 (2%)
- Missed one session accidentally	1 (2%)
- No transport	1 (2%)
- Sent to the wrong place	1 (2%)
- The location was misshared but the time was made up so content was missed	1 (2%)

N.B. Total percentage is below 100% because some respondents reported attending all training sessions.

3.3 After you received your training on how to deliver First Thing Music, how prepared did you feel to deliver the programme?

Response	Number (%)
I felt somewhat prepared	26 (54%)
I felt very prepared	20 (42%)
I felt neither prepared nor unprepared	2 (4%)

4.1 How often did you deliver First Thing Music sessions in your classroom?

Response	Number (%)

Daily	20 (42%)
Four times per week	15 (31%)
Three times per week	9 (19%)
Two times per week	2 (4%)
Less than once per week	1 (2%)
Don’t know	1 (2%)

4.2 On an average day, when did you carry out First Thing Music with your class?

Response	Number (%)
Early afternoon (12pm-2pm)	18 (38%)
Early morning (8am-10am)	16 (33%)
Late morning (10am-12pm)	10 (21%)
Late afternoon (2pm-4pm)	4 (8%)

5.1 To what extent do you agree or disagree with the following statement: I understand the purpose of First Thing Music

Response	Number (%)

Strongly agree	22 (46%)
Agree	22 (46%)
Neither agree nor disagree	3 (6%)
Disgaree	1 (2%)

5.2 To what extent do you agree or disagree with the following statement: I can deliver basic music exercises e.g. Ickle Ockle

Response	**Number (%)**
Strongly agree	27 (56%)
Agree	21 (44%)

5.3 To what extent do you agree or disagree with the following statement: I can keep the children engaged in the exercises

Response	**Number (%)**
Agree	30 (63%)
Strongly agree	18 (38%)

5.4 To what extent do you agree or disagree with the following statement: I can keep the children engaged in the exercisesI can structure First Thing Music sessions

Response	**Number (%)**
Agree	30 (63%)

Strongly agree	15 (31%)
Neither agree nor disagree	3 (6%)

5.5 To what extent do you agree or disagree with the following statement:I have all the materials required to deliver First Thing Music

Response	Number (%)
Strongly agree	23 (48%)
Agree	22 (46%)
Neither agree nor disagree	2 (4%)
Disagree	1 (2%)

6.1 What kind of impact, if any, do you think First Thing Music has had on pupils' music skills?

Response	Number (%)
Very positive impact	27 (56%)
Somewhat positive impact	18 (38%)
Neither positive nor negative impact	2 (4%)
Don't know	1 (2%)

6.2 What kind of impact, if any, do you think First Thing Music has had on pupils' behaviour?

Response	Number (%)
Somewhat positive impact	30 (63%)
Neither positive nor negative impact	9 (19%)
Very positive impact	7 (15%)
Somewhat negative impact	1 (2%)
Don’t know	1 (2%)

6.3 What kind of impact, if any, do you think First Thing Music has had on pupils’ reading?

Response	Number (%)
Neither positive nor negative impact	24 (50%)
Somewhat positive impact	17 (35%)
Very positive impact	3 (6%)
Don’t know	3 (6%)
Somewhat negative impact	1 (2%)

6.4 What kind of impact, if any, do you think First Thing Music has had on pupils’ speech and language?

Response	Number (%)
Somewhat positive impact	27 (56%)

Very positive impact	10 (21%)
Neither positive nor negative impact	10 (21%)
Don’t know	1 (2%)

6.5 What kind of impact, if any, do you think First Thing Music has had on pupils’ social skills?

Response	Number (%)
Somewhat positive impact	28 (58%)
Very positive impact	17 (35%)
Neither positive nor negative impact	2 (4%)
Don’t know	1 (2%)

6.6 What kind of impact, if any, do you think First Thing Music has had on pupils’ creativity?

Response	Number (%)
Somewhat positive impact	28 (58%)
Very positive impact	13 (27%)
Neither positive nor negative impact	6 (13%)
Don’t know	1 (2%)

6.7 What kind of impact, if any, do you think First Thing Music has had on pupils’ engagement?

Response	Number (%)
Somewhat positive impact	22 (46%)
Very positive impact	20 (42%)
Neither positive nor negative impact	5 (10%)
Don't know	1 (2%)

6.8 What kind of impact, if any, do you think First Thing Music has had on pupils' communication skills?

Response	Number (%)
Somewhat positive impact	25 (52%)
Very positive impact	15 (31%)
Neither positive nor negative impact	7 (15%)
Don't know	1 (2%)

7.1 How satisfied, if at all, do you feel about the level of support received from your Senior Leadership Team?

Response	Number (%)
Very satisfied	25 (52%)
Somewhat satisfied	11 (23%)

Neither satisfied nor dissatisfied	11 (23%)
Somewhat dissatisfied	1 (2%)

7.2 You worked with a specialist to help deliver First Thing Music in schools. How satisfied or dissatisfied were you with this support?

Response	Number (%)
Very satisfied	34 (71%)
Somewhat satisfied	10 (21%)
Somewhat dissatisfied	3 (6%)
Neither satisfied nor dissatisfied	1 (2%)

8.1 Overall, how would you rate First Thing Music?

Response	Number (%)
Very good	29 (60%)
Good	16 (33%)
OK	3 (6%)

Appendix L2. Raw control condition survey data

Please note: some percentages may add up to slightly above or below 100% due to rounding

1.1 What is your role in the school

Response	Number (%)
Class teacher	46 (92%)
Head or deputy head teacher	3 (6%)
Other members of the SLT (e.g. Head of KS2)	1 (2%)

2.1 Over the last academic year on average how often has your class received whole-class music lessons?

Response	Number (%)
Once a week	18 (36%)
Every couple of weeks	15 (30%)
Once a month	6 (12%)
More than once a week	4 (8%)
Less than once a month	4 (8%)
Never	2 (4%)
Don’t know	1 (2%)

2.2 Please tick which of the following musical activities your class have taken part in this school year

Response	Number (%)
Singing songs	49 (98%)
Listen to recorded music	39 (78%)
Playing instruments	32 (64%)
Performing music in front of an audience	28 (56%)
Listening to live music	15 (30%)
School trips related to music	8 (16%)
Other (please specify):	
- Charanga	1 (2%)
- Music appreciation	1 (2%)
- Repeating rhythms	1 (2%)

3.1 If another Year 1 class in your school is taking part in the First Thing Music programme, has any learning from the programme been shared with you?

Response	Number (%)
No	46 (92%)

Yes	3 (6%)
Don’t know	1 (2%)

3.2 If yes, what learning has been shared?

Response	Number (%)
We have shared best practice, discussing what worked well and what didn't	1 (2%)
I have attended all training but not used any of the work with my class	1 (2%)
Went to training	1 (2%)

N.B. Total percentage is below 100% because only respondents who reported having shared learning completed this question.

3.3 Has this influenced any of the lessons you deliver to your class?

Response	Number (%)
No	2 (4%)
Yes	1 (2%)

N.B. Total percentage is below 100% because only respondents who reported having shared learning completed this question.

3.4 If yes, can you describe the way in which it influenced your class?

Response	Number (%)
Improved children’s understanding of music	1 (2%)

N.B. Total percentage is below 100% because only respondents who reported having shared learning completed this question.

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

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

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

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

Education Endowment Foundation

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

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