

Testing the impact of high-quality music instruction on cognitive skills and academic attainment

Proposal to the Education Endowment Foundation for circulation, January 2013

1.1 Impact evaluation

This evaluation will test the effectiveness of teaching young children to sing or play a musical instrument on their cognitive development and educational attainment. A three arm (two treatment groups and a control group) individual-level randomised control trial research design is proposed. Year 2 pupils will be randomly allocated to two treatment groups and to an 'active' control group in approximately 15 primary schools which will be in disadvantaged areas. Treatment Group 1 will be taught to sing using the 'Kodaly Method', while Treatment Group 2 will be taught to play a string instrument. An 'active' control group will participate in drama lessons. Tuition will be delivered in small groups which we assume will consist of roughly ten pupils each – a point discussed further below. Steps will be taken to ensure that the quality of instruction across groups is uniform. The interventions will be delivered over one school year.

Due to the absence of an 'inactive' control group, comparisons will be made between two forms of musical tuition and drama. The use of drama as an active control helps counter *some* of the potential consequences of 'resentful demoralisation' and 'compensatory rivalry' among human subjects when knowledge of their allocation to a 'non-active' control group becomes known, as well as alleviating problems associated with 'novelty bias' (Hawthorne Effects). An active control also makes treatment contrasts explicit through providing the controls with a specified 'activity'. However, it is still possible that parents may compensate for allocation to a less preferred treatment by acquiring preferred treatments independently as a consequence of the study.

Information from Creative Futures, IoE and EEF suggest that the music tuition (singing and string instrument lessons) would be expected to improve the numerical and reading skills of participating children and that these outcomes would not be expected to be observed in the drama lesson control group. The current design means that any impact measured can only be relative to the active control drama group.

Another potential limitation of the proposed design is that children assigned to a given treatment (for example to learn a stringed instrument) may influence the outcomes of others in their class assigned to alternative treatments (for example to control conditions). This would be a violation of the assumption of non-interference between subjects; though we suspect, such effects might be non-trivial. The problem of interference between subjects is one motivating factor behind the choice of a cluster or multi-level randomised design; these approaches have, however, been ruled out in this case. Teacher/tutor effects will also need to be given consideration.

Research questions

Figure 1 provides an illustration of the design. The effects of the treatments can be estimated by comparing average outcomes between the three research groups at post-test. The research design permits a range of effects to be identified based on the following comparisons:

- Comparing Treatment 1 to Treatment 2 –the effects of learning to sing relative to learning to play a musical instrument
- Treatment 1 to Control –the effects of learning to sing
- Treatment 2 to Control –the effects of learning a stringed instrument
- Treatment 1 and Treatment 2 to control –the effects of musical tuition

This final comparison will only be made if no statistically significant difference in outcomes is observed from the comparison at bullet one above.

Proposed approach to randomisation & collection of pre-intervention test data

As Figure 1 illustrates, subsequent to school recruitment we will enumerate children in the target group at each participating school, clarifying who is eligible for the trial. This will be done twice for the target cohort, towards the end of Year 1 (May-June 2013) and in September 2013 at the commencement of Year 2. It is estimated that some 900 pupils in Year 2 will be targeted across 19 schools – on average 60 pupils in each school consisting of two classes or forms. Parental consent will be required for children to participate in the research; to be randomised into groups and to receive the music/drama intervention. Problems in obtaining consent may reduce the available sample size and rates of consent may vary across schools. We will obtain consent from parents prior to pre-test and randomisation using an opt-out approach co-ordinated with Creative Futures and IoE. Prior to the pre-test Creative Futures will collect background pupil information directly from schools and provide information to NatCen about the participating schools. At pre-test, outcome measures and simple demographics from each child for whom consent has been obtained will be collected in the school setting. Early discussions suggest that the music lessons (learning an instrument or learning to sing) would be expected to improve children's numeracy and reading skills (over and above normal development in these areas that would be expected to occur during the intervention). We will use standardised tests to measure children's reading and maths abilities. The reading and mathematics modules of the Performance Indicators in Primary Schools (PIPS) assessments produced by the Centre for Evaluation and Monitoring (CEM) will be used to test participating children. Each test module takes 30 minutes to administer and wherever possible children should complete the modules over two separate days. Tests will be administered in groups of 5/6 children by a NatCen interviewer.

NatCen interviewers will administer the tests in the school setting, in groups of 5/6 children. NatCen interviewers are experienced in using tests such as the BAS and working in schools but they would nevertheless receive a day of briefing and training prior to fieldwork. However, the PIPS assessments have not been used by NatCen on any previous studies so we will carry out a small pilot to test the procedures fully with interviewers prior to the main pre-testing. Interviewers will be responsible for setting up appointments in liaison with school staff. The following background information will be collected from teachers about each child at the enumeration stage: Unique Pupil Number, gender, date of birth, eligibility for Free School Meals (FSM), ethnicity and Special Educational Needs (SEN) status¹. This background information will help identify some important subgroups for analysis. We will also explore the possibility of linking study records to the National Pupil Database (NPD) which would enrich the data further for this analysis.

In order to make cooperation with the study more feasible for schools, we will obtain parental consent (via an opt-out procedure) and collect pre-tests at the end of Year 1 (hence enumeration at this point) – in this way only consent, pre-tests and the collection of background information for new starters needs be obtained at the beginning of Year 2 (where enumeration takes place again in order to identify leavers and new starters). This strategy should help reduce any delay in beginning the intervention, whilst consent and pre-test data are collected, making timetabling for the school more practical.

Thus in the first week of the new term (September 2013) random allocation will take place. We will stratify randomisation by form or class creating three groups of children in each form so that each child has an equal chance of allocation to the three arms of the study (other stratifiers will also be considered). If average class sizes are around 30, stratification will create three groups of roughly ten children in each form. Such an approach should help make the study more manageable for schools. Random allocation will be undertaken by NatCen statisticians.

Collection of post-intervention test data

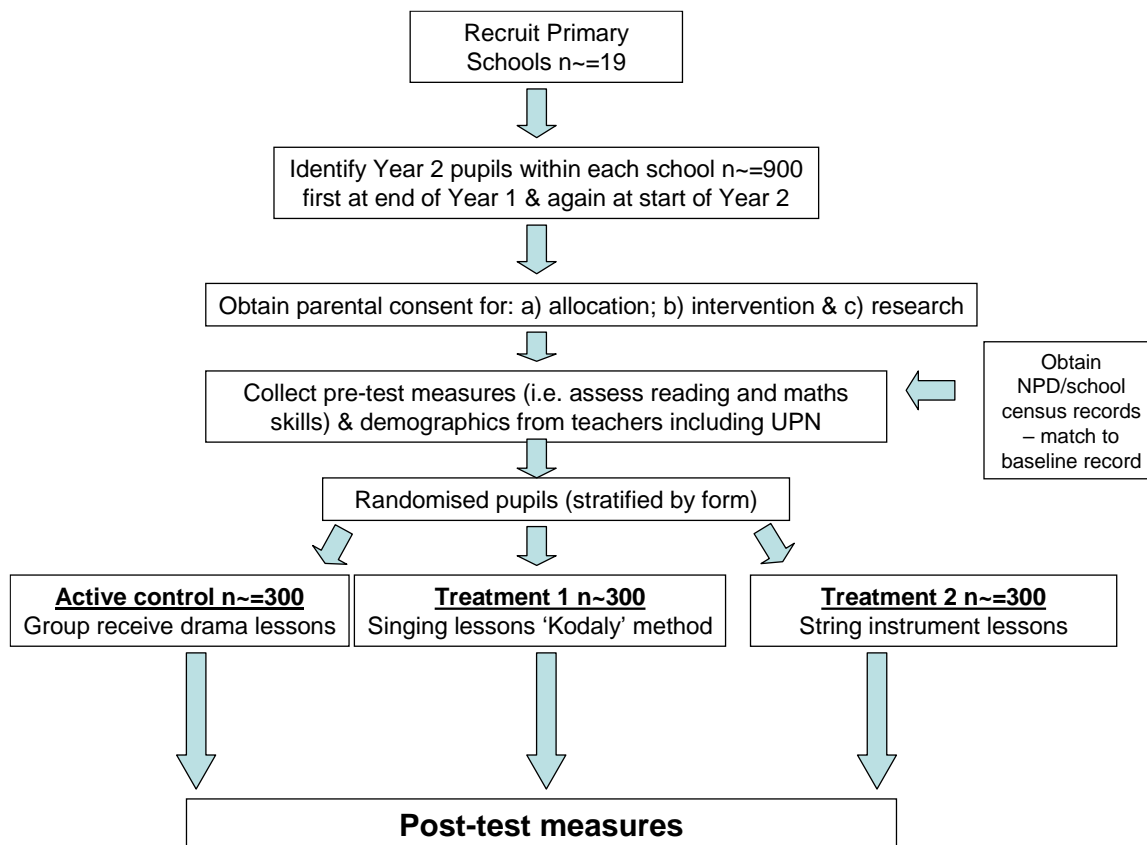
We will collect post-test data from all those pupils assigned to the three arms of the study at the end of the summer term 2014. Again, these data will be collected within the school setting and where possible, by the same NatCen interviewers who conducted the pre-test. The tests will be 'blinded' meaning that the interviewer will not know which intervention the pupil received. Loss to follow-up of study subjects will be minimised through positive engagement with schools, drawing on our experience of working with schools as research participants. This would include maintaining a project page hosted on the NatCen website with material of relevance to schools, teachers and parents and timely communication with schools by letter and email. Contact details for NatCen would be provided to schools and parents in case of queries or concerns.

¹ The school agreement document and opt-out letter sent to parents will include details about this data collection and reassurances about data confidentiality.

Monitoring random allocation

One potential problem is that teachers/practitioners may for a variety of reasons subvert randomisation or administrative errors might occur, leading to crossovers (children joining arms of the study they were not assigned to). We will monitor attendance at tuition groups closely by asking tutors to record attendance at the beginning of each tuition group and to forward these records to our research team on a frequent basis so that we can monitor attendance and compliance with randomisation.

Figure 1: Overview of Research Design



Impact analysis

Due to random allocation, the three study groups should be well-balanced at randomisation. Differential loss to follow-up may, however, occur during the course of the study and once post-test data are available we will need to check the sample to ensure that it remains well-balanced on pre-test/baseline measures. If this proves not to be the case, we will need to use regression methods more extensively or other non-experimental analyses to control for any observed biases.

Assuming that randomisation produces three study groups that are well-balanced, the **'primary'** impact analysis will compare average numerical skills and reading attainment (main outcome measures) at post-test in the three groups, on an intention to treat basis. This will be achieved by estimating a series of multiple regression models controlling for pre-test measures, stratification and possibly other factors.

We also propose to conduct some **'secondary'** analyses of differences across subgroups. Such analyses will be greatly enhanced with background demographic information (and more so if the NPD data is matched onto our study records) and therefore add to the items measured at pre-test/baseline. For example, it might be possible to explore whether impact varies by SEN status, ethnicity and eligibility for free school meals, sample sizes permitting. Any subgroup analysis will be clearly specified in advance after discussion with the research partners and make clear reference to theoretical considerations.

Due to the number of research groups, outcome measures and subgroup analyses proposed, a large number of statistical tests could be performed which might lead to the charge of ‘data dredging’. The chance of incorrectly rejecting null hypotheses where no true effect is present is also raised in such circumstances. We propose to counteract this problem by limiting the number of statistical tests conducted (in agreement with the partners) and specifying clearly in advance the types of tests to be performed and the covariates to be used in regression models. This will involve establishing *a priori* with the research partners a small number of ‘**primary**’ outcomes that will be the focus of analyses as well as identifying a range of further outcomes considered ‘**secondary**’. For example, we are likely to consider subgroup analysis as ‘**secondary**’.

Minimum detectable effect sizes

For this study we define a minimum detectable effect size² as the smallest effect size that if true will yield an estimate of impact statistically significant at the 5 percent level with 80 percent power. The estimates below are based on a comparison of outcomes between one treatment arm and another, and between the combined treatment groups and ‘active’ control group. A loss to follow-up rate of five percent is assumed along with no differential loss across the study groups³. Moreover, we assume that impacts are estimated using regression adjustment and that two-tailed statistical tests will be performed.

We estimate that for subgroup analyses based on comparisons between two groups containing approximately 100 children in each, an effect size of 0.4 standard deviations will be detectable at 5 percent statistical significant and 80 percent power. This assumes an unadjusted comparison.

Table 1: Minimum detectable effect sizes – whole sample estimates

	Comparison of one study arm with another	Comparison of treatment groups 1 and 2 with the active control group
Total assumed sample size	570	850
Unadjusted	0.23	0.20
.20*	0.21	0.18
.40*	0.18	0.16
.60*	0.15	0.13
* proportion of residual variance explained by covariates		

1.2 Process evaluation

The primary aim of the process evaluation would be to provide a detailed understanding of how the intervention programme is delivered in practice and to identify which elements are critical to it’s success. This evidence will be used to help understand and explain the impact results as well as to inform decisions about whether and how to scale up or roll out the Creative Futures programme more widely. This component of the evaluation will investigate key features of the music tuition, singing intervention and drama lessons including:

- the aims and delivery model of the stringed instrument and singing tuition programmes;
- the design and delivery of the training for tutors;
- variation in the delivery of the music tuition and the extent of fidelity to the programme model;
- who the tuition seems to work best for and the critical features;
- perceptions of the value, role and efficacy of the tuition;
- the content and delivery of the drama tuition; and
- other music tuition that may be operating in the schools to which the intervention or control pupils may be exposed, e.g. class music, school choirs and instrumental groups, music lessons.

We propose three methods to gather information about the delivery of the music tuition, singing and drama lessons that enable breadth and depth of coverage: analysis and monitoring of participating pupils, a short survey administered to all participating schools and qualitative data gathering through a school case study approach.

² By effect size we refer to the standardised mean difference

³ We have no made an estimate of the likely loss to the study resulting from an inability to secure parental consent

Profiling and monitoring participating pupils

Once the pupils are fully enumerated in September 2013, we will conduct some profiling analysis and then monitor pupil attendance at the intervention programme to which they are randomly allocated. This element of the study will involve the following activities:

- Matching the pupils to the National Pupil Database (if the DfE give permission for linking without parental opt-in consent);
- Contextualising pupils and school characteristics within the local authority and country as a whole to enable an assessment of the extent to which findings are generalisable;
- Using registers collated by Creative Futures to check that pupils are attending the intended programme consistently.

Profiling participating schools

The profiling exercise would involve gathering information on school characteristics and the implementation of the Creative Futures & IoE programme. The emphasis of this element of the process evaluation is on breadth, gathering top level information from all participating schools in order to understand programme delivery and variation. All 19 schools taking part in the trial would be invited to complete a short online survey to cover any required data not already held by the project team. The survey will be set up using Survey Monkey and will include a series of semi structured questions which will collect factual information about:

- The delivery of the sessions: group size, length and frequency of lessons and where tuition takes place;
- Whether the same tutor works with pupils throughout the year;
- Procedures for determining that the random allocation of pupils is maintained;
- Systems for encouraging engagement and perseverance;
- Qualifications, experience and training received by tutors;
- Other opportunities for music and performing arts offered within the school;
- Any information on school context that isn't available on Edubase/NPD.

The analysis of profiling returns will be used to identify any support required by schools to maintain the fidelity to the RCT model and to develop a typology of schools to better understand the likely heterogeneity of programme implementation in both the treatment and control groups across schools.

School case studies

The typology developed from the profiling stage will be used to purposively select 5 school case studies for follow-up interviews and observations, with the selection ensuring a range in the characteristics of schools that may be relevant for the trial.

We will carry out up to four 30-45 minute interviews in each of the case study schools with the key members of staff involved in the delivery of the music, singing and drama programmes and with a member of staff who has oversight of music tuition within the school. The interviews will investigate staff views on delivering the programmes, collect details of how the sessions are delivered and staff views of how well it is working and what is critical to its success. Interviews will also explore what types of music provision are available to children in the control group and explore with staff the potential for spill over effects from the intervention to the control group.

The interviews will be digitally recorded with the respondents' permission to generate data of sufficient quality for detailed and rigorous analysis and to prevent selective reporting. It also enables the interviewer to concentrate on the respondent, picking up essential non-verbal cues and engaging fully in exploratory and responsive questioning. This data will be analysed using Framework, a systematic approach to qualitative data management developed by NatCen Social Research and now widely used in social policy research.

Additionally, an observation of two music sessions (one of a stringed instrument and one of a singing lesson) would be carried out to provide evidence on the extent to which the pupils are engaged, the format of the session, the nature of the tuition, the process for recording who attends, the length of the session and any encouragement for practice before the next session. An interviewer will visit each school once, mid-way through the academic year to carry out observations.

1.3 Project team

The project would be located in the Children & Young People Team at NatCen. The trial manager would be Cheryl Lloyd (Research Director), and the impact evaluation project team would include Gemma Lewis (Senior Researcher) and Natasha Reilly (Researcher) both of whom have experience of quantitative research in schools. The process evaluation would be managed by Amy Skipp (Research Director) working with Meg Callanan (Senior Researcher) and Natasha Reilly (Researcher). The researchers would work closely with other departments and specialists at NatCen. Stephen Morris (Head of Evaluation), Kevin Pickering (Head of Statistics) and David Hussey (Research Director Statistician) would be closely involved in the evaluation design, selection, randomisation and analysis, bringing their experience of running experiments and randomisation. The Operations Department would coordinate the contact with schools, testing and data entry processes.

Main risks to evaluation and mitigating actions: Analytical, procedural and managerial

Risk	Likelihood	Impact	Mitigation
Crossovers - children participating in treatments they were not allocated to receive, due to error/subversion	Medium	Medium	We propose to monitor receipt of treatment carefully to ensure the integrity of the study design is maintained and ensure that crossovers are minimised
Drop out – school may decide to leave the study thus reducing the study sample size	Low	High	We will draw on our experience of working closely with schools to ensure interest in and cooperation with the study is maintained, addressing practical concerns as necessary, and ensuring the benefits of continued participation are made clear.
Difficulty in obtaining parental consent – risks to sample size and under-representation of pupils with certain characteristics. Consent rates may vary across schools.	Low	Medium	We will work hard with schools and research partners to obtain consent from as many parents as possible, ensuring parents are aware of the benefits. If handled carefully, this consent should not be a significant problem – parents are used to being asked for their consent on a regular basis for a variety of school activities.
Variations in tutor quality by treatment – could lead to false conclusions regarding effectiveness of interventions.	Medium	High	We will explore variation through the survey and observations and will explore ways in which the analysis can be adjusted for tutor effects. We will ensure for each member of the study we record that tutor they were assigned to.
Loss to follow-up – could be a problem if overall sample loss is large (reducing absolute sample numbers)/patterns of loss differ between study groups.	Low	High	The best way to tackle loss to follow-up is through well-designed fieldwork procedures. Our approach to data collection minimises teacher burden, testing through CAPI maximises efficiency, and our communication procedures are of a high quality.
Treatment diffusion – music teachers in study schools may incorporate aspects of treatments into their routine practice potentially affecting control pupils.	Low/medium	Medium	As part of our general close liaisons with schools we plan to check for this. We also propose to interview school music teachers as part of the process evaluation.
Variations in treatment delivery – significant departures from specified treatments may affect interpretation of treatment effects/or reduce effectiveness.	Low/medium	High	This risk will be controlled by study partners who will deliver treatments. The process evaluation will explore variation in treatment delivery across schools and consider the implications for study results. There may also be potential for feedback/corrective action.
NPD data access denial	Low	Medium/high	Were we unable to negotiate access to NPD records, we propose to work closely with schools in order to gain access to school level records used to compile the NPD
Lack of parent level data	High	Low/medium	Ideally we would like to collect data from parents but this would add substantially to costs and risk significant non-response. Lack of parent-level data will be ameliorated by accessing NPD data prior to randomisation to supplement pre-test/baseline measures.
Difficulty scheduling case study visits.	Medium	Low	We will schedule case study visits in the North on consecutive days so as to allow flexibility in rescheduling if necessary. Interviews can be completed by phone.
Changes to project specification affecting timetable/costs	Medium	Medium	We have allocated sufficient resources to the set-up stage for the design to be clarified in detail allowing us to revise costs/timing where necessary and provide a robust estimate.
Poor project management	Low	Medium	We have proposed a strong team experienced in managing complex evaluations. We will monitor progress closely and identify areas of concern early.
Staff illness / unavailability / turnover	Low	Low	We forward-plan research capacity, and have a sufficient number of experienced staff members. Our procedures ensure that decisions and progress are fully documented.
Loss of or damage to data	Low	High	NatCen has high quality data security procedures with which team members are experienced. EEF will be notified of any breaches and contingency plans put in place.

Timeline

Activity	Responsibility	2013												2014											
		J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D
Project set-up: finalise design, tests, budget, contract, meet project team.	All																								
School recruitment	Creative Futures (CF)	■	■																						
Assist with selection of schools if required and make contact	NatCen	■	■																						
Enumerate pupils in target group for intervention	NatCen						■	■		■															
Pilot (3 schools)	NatCen					■																			
Brief interviewers, prepare test materials and schedule visits	NatCen					■	■																		
Pre-tests	NatCen						■	■		■															
Data entry	NatCen								■	■															
Match pupil details to NPD (if permission given)	NatCen									■	■														
Random allocation of pupils to groups	NatCen									■	■														
Intervention and monitoring	CF NatCen									■	■	■	■	■	■	■	■	■	■	■					
On-line survey	NatCen									■	■	■													
School case studies	NatCen													■	■	■									
Profiling analysis	NatCen													■	■	■									
Post-tests	NatCen																■	■							
Data preparation, analysis & reporting	NatCen																		■	■	■	■	■	■	