

Protocol: Evaluation of CfBT's 'Foreign Language Learning' programme – Institute of Education

Meg Wiggins (project leader), Dr. John Jerrim, Dr. Shirley Lawes, David Mallows, Helen Austerberry

This evaluation aims to measure the CfBT's Foreign Language Learning programme for effectiveness in raising English language attainment in primary school children, as well as the acceptability and potential sustainability of this programme.

Background

A new 'Languages Programme of Study' for English primary school pupils was issued in 2013 by the Department for Education, as part of the statutory guidance relating to the new National Curriculum¹. Within this new scheme, beginning in the 2014/15 school year, all Key Stage 2 children will be expected to be taught a foreign language, providing the groundwork for more in-depth study of a modern foreign language in Key Stage 3.

Two previous large scale studies in the US indicated that the introduction of the teaching of a foreign language had improved the English language skills of those who were taught an additional language versus a matches sample of those with no foreign language teaching^{2 3}. No large-scale experimental studies from the UK address this area. The CfBT intervention being evaluated includes lesson plans for the teaching of French to Key Stage 2 children, in line with the attainment criteria laid out in the new National Curriculum. Additionally, the intervention provides lesson plans that provide specific links between French and English literacy. This evaluation was designed to allow a comparison of children exposed to this language intervention with those receiving the standard language provision (if any) provided by their school, prior to the introduction of the new National Curriculum programme in September 2014. The evaluation was pragmatically designed to fit in with this very tight timetable for the recruitment of schools, intervention training and delivery and evaluation testing.

The Intervention

¹ Department for Education (2103) Languages programmes of study: key stage 2. National curriculum in England.

² Rafferty, E. A. (1986) Second Language Study and Basic Skills in Louisiana. Baton Rouge: Louisiana State Dept of Education.

³ Taylor-Ward, C. (2003) The relationship between elementary school foreign language study in grades three through five and academic achievement on the Iowa tests of basic skills and the fourth-grade Louisiana educational assessment programme. Dissertation submitted to the Graduate Faculty of the Louisiana State University.

The CfBT Foreign Language Learning programme consists of the following key components:

- 1) The programme curriculum**, which is based on the new National Curriculum Programmes of Study for English and Modern Foreign Languages, and consists of three half term units of work. There are three units which have been designed for use with Year 3 classes; three for use with Year 4 classes. Each unit includes:
 - Six step by step lesson plans for teaching French – each lesson is designed to last approximately 45 minutes.
 - Follow up activities linking the learning of French to English literacy. The follow up activities focus on three areas – extending vocabulary, phonology and grammar/punctuation. An additional teaching time of between 15 – 30 minutes per week is estimated to be needed to cover these follow up activities.
- 2) Teacher training in the programme:** Year 3 and 4 class teachers will receive three one-day training sessions, held in groups with teachers from other schools in the study. The programme developers from CfBT will lead the sessions, which will train teachers in the approach, and provide specific instruction about the unit to be taught in the subsequent half term to each year group.

Evaluation Design

To evaluate the CfBT's Foreign Language Learning project we will conduct an impact study that has the capacity to show effectiveness, using a cluster randomised trial. This will be combined with an integrated 'light touch' process evaluation that has the scope to provide valuable insight into why the intervention has, or has not, had the intended impact on attainment, as well as explaining the variation in use and implementation of the Foreign Language Learning programme by different teachers.

The Impact Evaluation

Given the restraints on timing, our pragmatic impact evaluation will include the following components:

- 1. A cluster randomised design**, with randomisation of treatment at the class level within approximately 30 primary schools. Schools with two-form entry will be recruited, and in each of their Year 3 and 4 groups, the year group classes would be considered matched pairs.
- 2. Status quo control.** The comparison will be between pupils in those classes receiving CfBT's Foreign Language Learning programme versus those in classes receiving 'status quo' control of current school teaching in literacy and foreign languages.

3. **No baseline testing or surveying**, instead using as a proxy baseline NPD data from teacher assigned Key Stage 1 SATs scores.
4. **End of programme testing** carried out by researchers in schools.

Further detail of each of these components is provided below.

(i) Research question – Impact Evaluation

The key research question to be answered by the impact evaluation is:

- Will pupils in classes allocated to the CfBT Foreign Language Learning intervention perform better on English grammar and vocabulary tests than the pupils who had existing school provision of language teaching?

(ii) Level at which treatment will be randomly assigned

Randomisation will be carried out at the class level, given that this intervention is designed to be used with a full class of 30 pupils. Schools with at least two-form entry will be recruited, and in each of their Year 3 and 4 groups, the year group classes would be considered matched pairs with one receiving intervention and the other acting as the control. In schools with three-form entry, one year group class will be allocated to intervention and two to control.

The selection of class level randomisation, rather than school level randomisation, was made because of the limited time for school recruitment and the capacity of the intervention team. In addition to being a pragmatic design, this approach will provide more statistical power than some of the possible alternatives (e.g. school-level intervention); and because each school would receive some intervention, this design might reduce the attrition commonly seen in control schools. This approach also has one major potential limitation: parents might find it inequitable to have some year group children being provided this foreign language programme while others do not receive it. However, the brevity of this intervention (18 weeks), the unproven benefit of the programme, the ‘business as usual’ control (which will mean foreign language teaching for some control classes) and the fact that all pupils will receive foreign language teaching in September 2014 as part of the curriculum, suggest that this randomisation has an ethical basis. Other possible challenges of this design include the fact that some schools might find this organisationally difficult given joint curriculum planning within year groups; and there is a risk of some diffusion of the intervention between classes, especially if there is existing language teaching at the school. These limitations will be discussed at the training for the programme, to ensure as much co-operation as possible from teachers for the short duration of this programme.

(iii) Outcome measures of interest

The CfBT programme team highlighted three possible outcome measures relating to pupils' attainment in English literacy: vocabulary, phonology and grammar, which they hypothesised would be impacted by participation in their Foreign Language Learning intervention. Our evaluation will measure two of these three: vocabulary and grammar. Phonology will not be included as an outcome given the challenges in testing this domain and concerns about over-burdening schools.

We will create a single primary outcome that is a scaled average of scores from tests in vocabulary and grammar. Analysis on outcomes will be by intention to treat in the first instance. Secondary analysis will explore the differences in effect based on dose of intervention, as well as on the two domains of vocabulary and grammar separately. A subgroup analysis will be performed by Free School Meal (FSM) status and gender, via an interaction with treatment status.

(iv) Testing

We do not propose to conduct a baseline test and hence a direct comparison of changes in outcomes for treatment and control groups will not be possible. However, our intention will be to gain some information on pre-treatment ability in the form of teacher assessed KS1 and Foundation Stage Profile scores, via a link to the National Pupil Database. Our reasoning for this choice is because of two pragmatic factors: i) the tight timetable of this project suggests that there will already be increased pressure and burden on participating schools; asking for additional changes to timetables to accommodate two rounds of testing may lead to greater attrition and reduce the amount of time to deliver the intervention; ii) the cost-effectiveness of baseline testing when compared with using existing KS1 SATs scores as a baseline. The absence of a baseline test reduces statistical power for a given sample size but the randomised assignment of treatment will continue to produce unbiased estimates of the treatment effect. The KS1 SATs scores can be used as a covariate in regression analysis of the test outcome data, thus reducing the possibility of any confounding and sustaining power. We require Unique Pupil Numbers to be provided to us by participating schools and for them to sign the consent documentation allowing the project to link to their data to NPD.

We propose to use two tests to measure two domains of English literacy: vocabulary and grammar. We will use a fieldworker administered test, the British Picture Vocabulary Scale, to measure development in vocabulary. This will be undertaken with all intervention and control pupils in each class. A self complete test that includes components of grammar and general literacy, the short form Progress in English test from GL Assessments, will be undertaken by a whole class together, overseen by a researcher, to minimise burden on classroom teachers, ensure completion and safe transit of test papers. The testing will be carried out shortly after the intervention is complete. Parents and young people will be given information prior to the testing and opt out consent will be obtained from both groups.

(v) Appropriate control 'treatment'

In this study we proposed to have the control classes receive their school's usual, status-quo, curriculum for literacy and languages. There will be no alternative programme for these classes. This design will minimise organisational changes for schools. This may mean that some control classes are being taught a foreign language (including French), sometimes by specialist language teachers, but these classes would not have the CfBT training, approach or materials. This design will allow us to answer the question 'Does this intervention boost performance versus what is currently happening?' .

(vi) Sample size and power

The required sample size depends upon a number of factors, including the expected effect size of the treatment, the choice of level at which to randomise treatment, the extent of clustering in the data (the degree of correlation of pupil abilities within schools), and school size. Evidence from a large trial in the US indicated effect sizes ranging between 0.2 and 0.4 on key outcomes of vocabulary and grammar⁴ for children who received foreign language instruction when compared to those who had not had this. Given that this is a relatively short intervention, and that the control group will include some language teaching, we feel a more conservative estimate of overall effect would be more realistic when producing a power calculation for this study, and aimed for 0.20.

We have estimated power calculations based on class level interventions. At the class level, we have assumed that there is an intra-cluster (between class) correlation of about 0.05, a fixed year size of 30 pupils per class and a correlation between KS1 and the outcome measures of 0.60. This would require approximately 60 clusters (classes) or roughly 30 schools (assuming only 2 form entry schools were recruited) to detect an effect of 0.20. The main risk to this calculation is that the between class correlation is higher than expected (if we assumed a between class correlation of 0.10 instead of 0.05 one would require 90 classes across 45 schools).

CfBT will have responsibility for recruiting the schools. Given the tight timetable for recruitment, rather than defining the sample from which to recruit, CfBT will note all schools approached and their responses, so that we can define the characteristics of those that participated and those that refused. The eligibility criteria for schools to be recruited include: primary schools with a minimum of two-form entry; willingness of the school, and teachers in years 3 and/or 4, to allow within school randomization by class for training and delivery of the intervention.

(vii) Analysis

Following recruitment of schools we will assess external validity by analysing the pattern of school response among sampled schools. To check whether randomization has indeed balanced observed potential confounding factors, we will be able to compare characteristics

⁴ Rafferty, E. A. (1986) Second Language Study and Basic Skills in Louisiana. Baton Rouge: Louisiana State Dept of Education.

of children in treatment and control groups through the additional information included in the NPD (e.g. FSM status, ethnicity, Foundation Stage Profile and KS1 scores). We will also be able to include these variables into our analysis to (i) further reduce the possibility of any confounding and (ii) limit the decrease in statistical power from not having a baseline test.

As noted above, analysis on our primary outcome of vocabulary and grammar will be carried out initially on an intention to treat basis, followed by secondary analysis based on intervention dose. Additional secondary analysis will explore the differences in effect within the two domains of vocabulary and grammar separately. A sub-group analysis will be performed by Free School Meal (FSM) status and gender, via an interaction with treatment status. Our comparative statistical analyses will be carried out using Stata.

The Process Evaluation

Our light-touch process evaluation will work alongside the impact evaluation to help understand the presence or absence of treatment effects. It will assess the feasibility and acceptability of the intervention, in order to determine the potential for scalability, should the intervention prove to be effective. The process evaluation will address a variety of questions listed below.

(i) Process Evaluation - Research Questions

- How feasible and acceptable is it for teachers to teach foreign languages, using the intended evidence based programme, at KS2? How faithful were they able to be to the original programme plan? What are their views on continuing with the programme beyond the study period?
- How feasible and acceptable do the different 'types' of teachers (class teachers; specialist language teachers) feel it is for primary school children to be taught languages using the programme?
- How do children appear to engage with the language teaching using the selected teaching approaches? How does this vary by different subgroups (e.g. boys vs girls, ethnic groups)?
- What are staff perceptions of the current and possibly sustained impact of the intervention on children's English language attainment and specifically on vocabulary, phonology and grammar? How do they think it affects different sub groups (e.g. more or less disadvantaged children)? What are their perceptions of facilitators and barriers to impact?

(ii) Components of the process evaluation

1. Observation of the training sessions for teachers. One researcher will carry out non-participant observation at each of three modules of teacher training in order to describe the

training process, understand what is required for achievement of fidelity and assess feasibility and satisfaction from the perspectives of trainers and trainees.

2. Record of unit completion Throughout the 18 week programme, class teachers will be asked to record, on a simple record sheet, which parts of each weekly lesson that they delivered to their class, and provide an assessment of the ease or difficulty of doing these parts.

3. Online survey of teachers. Towards the end of the intervention period, teachers will be asked to complete this survey. It will be designed to capture data from class teachers without a specific subject specialist role, as well as staff with foreign language or literacy specialist/co-ordinator responsibilities. Participants will be asked to provide relevant background information on their class and on themselves, including any experience of foreign language learning/teaching prior to the study period. They will also be asked for their views on taking part in the study. Teachers will be asked questions on their experience of the training for, and the delivery of, the intervention, including their views on fidelity to original intervention plans and current and future acceptability and feasibility. The questionnaire will be used to gather data regarding the extent of language teaching in control classes. For non-responders we will send an email reminder and if non-response persists we will offer an alternative of a postal questionnaire or completion over the telephone.

4. Observations of foreign language and literacy lessons. One researcher will carry out non-participant observation of at least one language and one literacy lesson in four schools in order to describe the delivery of the intervention, to acquire insights into fidelity and to observe any transfer of the approach across-subjects. This will also provide some data on how the intervention is received by the children. The four schools will be purposively selected using selection criteria such as geographic region, previous school attainment and percentage of FSM. The researcher will spend a day in each of the 4 schools.

5. Interviews with teachers (class teachers and those with subject specialism responsibility). These interviews will be carried out in the schools where the observations have taken place. Ideally they will be conducted face to face during the one day school visit – but if this is not possible a date will be set to complete this over the telephone. The interview will provide the opportunity to clarify issues of interest arising from other sources of data. Interviews will be digitally recorded. Notes will be taken during the interviews and typed up afterwards. This will be supplemented with selective transcription to ensure accuracy of quotes. These data collection decisions are informed by our previous experience and based on an awareness of the high cost of transcribing interviews.

6. Process questions as part of intervention children's testing. As part of the testing process, we will survey children in both intervention and control classes with a few additional questions regarding the learning of foreign languages at school. This will help determine engagement with and acceptability of the intervention, as well as a comparison

with engagement with status quo language teaching. Consent procedures for these questions will be incorporated in those for the testing process as a whole.

(iii) Analysis

We will use Framework Analysis for the analysis of the qualitative data. This involves the construction of frameworks based on key themes that answer the main research questions. This method affords the possibility of exploring the data by both theme and respondent-type, so we might better describe and explain the data through the identification of patterns and associations across and between themes and types of respondents.

We will carry out descriptive statistical analyses of the teacher and pupil surveys. A range of strategies will be used to integrate the analysis, and the different types of process and impact data.

The Project Team

The project will be led by **Meg Wiggins**, Senior Research Officer in the Social Science Research Unit at the Institute of Education. Meg has over 20 years' experience leading large scale mixed-method research, including RCTs of complex social interventions. Meg is a very experienced project manager, having directed integrated impact and process evaluations in the health, education and community sectors. The randomization and quantitative analysis will be undertaken by **Dr John Jerrim**, a Lecturer in Economics and Social Statistics at the IoE. John has experience of evaluating several EEF randomized controlled trials, including 'Chess in Schools' and 'Maths Mastery' primary and secondary school interventions. **Dr Shirley Lawes** is Subject Leader for PGCE Modern Foreign Languages at the Institute of Education, University of London. Shirley has a background of experience in teaching French in Secondary schools, Further/Vocational Education, Adult Education and industry and for the past 18 years has worked in the university sector as a PGCE tutor and researcher. She has published widely on a number of aspects of foreign language teaching and learning, on teacher education and on broader educational issues. She was recently a member of the Ministerial Steering Group on Languages. **David Mallows** is a language teacher and teacher educator at the IoE. He directs national research and development projects in the areas of language, literacy and numeracy across all phases of education. His key research interests include teaching practice, initial teacher education and teacher learning. The process evaluation will be led by **Helen Austerberry** and assisted by **Anne Ingold** (both very experienced research officers in the Social Science Research Unit).

Risks

The main risk to this evaluation is the very tight timetable – for recruitment of schools, training of teachers, delivery of the intervention, and for evaluation testing. Additionally there are risks that the impact study design of randomisation by class will lead to diffusion between the intervention and control teachers. To minimize this risk, at the first training

days for teachers, the trainers will stress the importance of not sharing with control colleagues the details of the approach or the materials.

Timeline

September 2013 - December 2013	Schools recruited into trial; treatment and control schools assigned; treatment teachers attend 1 st phase training. (observation of training)
January 2014	Parents' letter distributed in all schools (control and intervention classes – opt out consent for testing). Trial intervention begins in years 3 and 4. Teachers record lesson completion
Feb 2014	Observation of second teacher training
March 2014	Select case study sites
April – May 2014	Observe final teacher training component
April 2014 - July 2014	Case study site visits for class observation, teacher interviews
July 2014	On-line teacher survey
June 2014	Additional information letter sent to parents about testing, allowing another opt out opportunity in week before test. Testing of children in Literacy components
April- June 2015	Data analysis; IoE write final EEF report