



Education
Endowment
Foundation

Talk for Literacy

Evaluation Report and Executive Summary

February 2015

Independent evaluators:

Ben Styles

Sally Bradshaw



Evidence for
Excellence in
Education

The Education Endowment Foundation (EEF)



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;
- 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 Trust (now part of Impetus-The Private Equity Foundation) 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 Sharples

Senior Researcher
Education Endowment Foundation
9th Floor, Millbank Tower
21-24 Millbank
SW1P 4QP

p: 02078021921

e: jonathan.sharples@eefoundation.org.uk

w: www.educationendowmentfoundation.org.uk

About the evaluator

The programme was independently evaluated by a team from the National Foundation for Educational Research. The lead evaluator was Dr Ben Styles. For the impact evaluation he was assisted by Sally Bradshaw and for the process evaluation by Eleanor Stevens, Rebecca Clarkson and Katherine Fowler.

Contact details:

Dr Ben Styles
The Mere
Upton Park
Slough
Berkshire
SL1 2DQ

p: 01753 637386

e: b.styles@nfer.ac.uk

Contents

Executive Summary4

Introduction6

Methodology8

Impact evaluation14

Process evaluation.....21

Conclusion.....28

Appendices31

Executive Summary

The project

The speaking and listening intervention evaluated was a combination of two programmes: the Vocabulary Enrichment Intervention Programme (VEIP) and the Narrative Intervention Programme (NIP). The former aims to teach children new words and to encourage the use of these words in speaking and writing. The latter aims to enhance the understanding and expression of narratives to develop speaking and listening skills. The dual intervention was used with pupils who needed extra support to improve their literacy (either not having reached Level 4 at Key Stage 2 or considered 'vulnerable' Level 4 readers). The aim was to enhance pupils' literacy by improving their vocabulary and narrative skills.

Five Teaching Assistants (TAs) and one teacher in the three secondary schools received training on the VEIP, the NIP and speech, language and communication difficulties. They delivered the VEIP to 118 Year 7 pupils in small groups (three to eight pupils) between September 2013 and December 2013 (approximately 12 teaching weeks) and the NIP between January 2014 and March 2014 (approximately 11 teaching weeks). Each intervention group had two lessons a week (40–60mins).

The programme was evaluated using a randomised controlled trial which compared the intervention to a 'business-as-usual' control group. The delivery and evaluation of this project was funded by the Education Endowment Foundation as one of 23 projects focused on literacy catch-up at the transition from primary to secondary school. It was one of four projects funded with a particular focus on reading comprehension.

Key conclusions

1. The speaking and listening intervention had a moderate impact on overall reading ability but this was not statistically significant (although was on the border of being so).
2. The intervention had a significant impact on pupils' ability in passage comprehension; an effect size of 0.25, equivalent to approximately 3 months of additional progress compared to control pupils.
3. The intervention had no significant impact on pupils' ability to complete written sentences or accurately recall spoken sentences.
4. The reasonably fast pace of delivery and necessarily selective approach to the programme materials (given the time available) may have limited the intervention's impact.

What impact did it have?

The main result of the trial was that the speaking and listening intervention had an effect size of 0.20; roughly equivalent to three months of extra progress in reading. However, the result was not statistically significant at the 5% level (although was on the border of being so). It did have a significant impact on one secondary outcome, passage comprehension, but did not have an impact on pupils' ability in sentence completion or recalling sentences.

Observations and interviews with the TAs indicated that pupils engaged with the lessons for the most part, although in some groups pupils' focus waned towards the end of each programme. In some cases, poor attendance resulted in some groups achieving limited coverage of the programme material (especially the VEIP), however there was no significant relationship between impact and the intervention 'dosage' (the time that a pupil spent in intervention lessons). The reasonably fast pace of

delivery and necessarily selective approach to the programme materials (given the time available) may have limited the intervention's impact. TAs reported that the intervention improved pupils' communication skills (in terms of articulation and confidence), rather than their reading ability. Those aspects of verbal communication were not formally assessed.

How secure is this finding?

Impact was assessed through a two-arm pupil-randomised controlled trial in three secondary schools: 235 Year 7 pupils were randomised either to receive the speaking and listening intervention or to be in the 'business as usual' control group.



The VEIP and the NIP had not previously been evaluated for their impact on reading comprehension. They had been trialled in a three-armed trial undertaken by the developer of the two programmes in 21 secondary schools in outer London areas from 2006. The developer reported a significant impact of the VEIP on vocabulary test scores and of the NIP on storytelling skills, although the full analysis and report has not yet been published. The present study represents the first evaluation of the speaking and listening intervention in terms of its impact on reading ability. As the delivery was overseen by the trial developers, it can be regarded as an efficacy trial. Efficacy trials seek to test evaluations in the best possible conditions to see if they hold promise. They do not indicate the extent to which the intervention will be effective in all schools since the participating schools are selected from one area, and the programme is closely managed by the developers.

The primary outcome was reading ability as assessed by scores from the GL Assessment *New Group Reading Test* (NGRT). The secondary outcomes were the two NGRT subscales of sentence completion and passage comprehension. Test administration was carried out by staff at the participating schools. Efforts were made to ensure that test administration was blind in every school although complete blindness, as if delivered externally, cannot be guaranteed. The test marking was carried out by GL Assessment and was therefore blind.

Ninety-one per cent of randomised pupils were included in the final analysis for both the NGRT and recalling sentences outcomes. Both control and intervention groups suffered similar levels of attrition and further analysis shows there is no evidence of bias in pupil characteristics between groups.

How much does it cost?

The cost of training and equipping two TAs to deliver the intervention to one cohort of 40 pupils each year for two years works out at £29 per pupil. This assumes a one-off training cost of £2,000 and provision of teacher resource packs at £87.50 each.

Group	No. of pupils	Effect size (95% confidence interval)	Estimated months' progress	Evidence strength*	Cost**
Intervention vs control (all pupils)	213	0.20 (-0.02, 0.42)	+3 months		£
Intervention vs control (FSM pupils)	60	0.33 (-0.06, 0.71)	+4 months		£
*For more information about evidence ratings, see Appendix B in the main evaluation report. Evidence ratings are not provided for sub-group analyses, which will always be less secure than overall findings.					
**For more information about cost ratings, see Appendix C in the main evaluation report.					

Introduction

Intervention

The speaking and listening intervention comprised two programmes delivered sequentially. The Vocabulary Enrichment Intervention Programme (VEIP) aims to teach children new words and encourage the use of these words in speaking and writing; the Narrative Intervention Programme (NIP) aims to enhance the understanding and expression of narratives to develop speaking and listening skills. These programmes were created by Professor Victoria Joffe (City University London) for children with communication and language impairment. In this trial the intervention was delivered in three schools to small groups of between three and eight Year 7 children who were considered to be vulnerable Level 4 or below at Key Stage 2 English (as indicated by a reading test administered at the end of Key Stage 2). Two teaching assistants (TAs) in each school delivered the intervention (with the exception that in one school one deliverer was a SEN teacher) during withdrawal or timetabled sessions. They taught the VEIP sessions for up to 12 weeks and then the NIP sessions for up to 11 weeks. Each group of pupils was intended to receive two sessions a week of between 45 and 60 minutes (the duration matched that of a normal lesson which varied between schools). In this report, the deliverers will be referred to as TAs.

The VEIP has two aims: to teach children new words, and to encourage them to use these words in speaking and writing. Its objectives include cultivating children's enthusiasm for words, developing their understanding of words in context, developing understanding of non-literal language, and teaching them specific strategies for learning and recalling new terms. Children are taught to recognise the structure of words (prefixes, suffixes, roots), are given a range of cueing techniques to aid retrieval, and are shown how to understand and learn new terms independently so they can continue to learn after the intervention has finished. The NIP is a practical language programme that focuses on enhancing the understanding and expression of stories through systematic explanation of the principles of good narrative (structure, mood, characterisation, and so on) identifying different types of narrative, and developing speaking and listening skills. The VEIP and the NIP seem to complement each other. The NIP sessions, for example, include revising concepts that are taught in the VEIP and applying techniques learned from the VEIP, such as using descriptive and figurative language, as pupils develop their storytelling skills. The programmes were chosen on the premise that pupils' narrow vocabulary, and weaknesses in recalling vocabulary and in expressing and understanding narrative, could contribute to their lower attainment in reading comprehension and in writing.

Background evidence

The two intervention programmes, the VEIP and the NIP, had been piloted as a study in 21 mainstream secondary schools in outer London boroughs. The study commenced in October 2006, and was funded by the Nuffield Foundation (Nuffield Foundation, no date). It involved 352 pupils aged 12 who had achieved Level 4c or below in Key Stage 2 English and who had speech, language and communication difficulties (SLCD). Eligible pupils were randomised to receive one of the VEIP, the NIP, aspects of both programmes, or business as usual. The intervention programmes were delivered by TAs and Learning Support Assistants (LSAs) with support from speech and language therapists. Pupils were intended to have three 45-minute sessions per week for six weeks.

The full analysis and findings from this trial have not yet been published, although the developer has published some headline findings.¹ The study reported significantly greater improvements in intervention group scores (compared to the control group) on a range of non-standardised tests

¹ Some headline results have been published by the developer in the form of a project newsletter (Joffe, 2011) and on the Nuffield Foundation project web page.

(including vocabulary tests and storytelling tests). Impacts on the test results were differential by trial arm: only the children in the VEIP group did better on the vocabulary tests, and only the children in the NIP group did better on the storytelling assessment. Children who received the combined intervention scored better than the control group on both vocabulary and storytelling (Joffe, 2011). Some of the trial schools continued and/or expanded use of the intervention following the trial, and a number of other schools in London have since adopted the programmes.

The Vocabulary Enrichment programme has also been trialled as a component of the Bolton Local Authority-led Vocabulary Enrichment Full Programme (NFER, 2014).

In the Nuffield Foundation-funded trial described above, pupils were excluded from randomisation and allocated to the treatment groups if they had English as a second language, Level 4b or above in Key Stage 2 English, or more general learning difficulties other than SLCD. In this trial a similar attainment exclusion was made but other exclusions were not. The present trial includes a broader range of pupils (not specifically those with specific language impairment). The trial has been designed as an efficacy trial with a small number of schools.

The programmes focus on two key skills targeted in the National Curriculum: storytelling (for instance, to promote complex use of language) and vocabulary enrichment (to promote literacy). In addition to preparing TAs to deliver the sessions, the training aims to improve their knowledge and awareness of speech, language and communication difficulties.

Evaluation objectives

The impact evaluation sought to answer the following research questions:

1. What is the impact of the speaking and listening intervention on reading ability?
2. Are improvements in attainment moderated by any of the following: prior attainment in reading, gender, whether a pupil receives the pupil premium, or the school they attend?

The purpose of the process evaluation was to assess the intervention in terms of fidelity to the programme intentions and scalability.

Project team

An assistant headteacher and a literacy leader at one of the participating schools chose the published speaking and listening programmes used in the intervention. The intervention was delivered by TAs (and one teacher) at the participating schools, supported by the developers. Paper-based tests were administered by school staff who were not involved in delivering the intervention and these were externally marked. The additional one-to-one tests were administered by staff contracted by NFER who were blind to the randomisation. The evaluation team at NFER was led by Dr Ben Styles; Sally Bradshaw assisted with the impact evaluation; Eleanor Stevens, Rebecca Clarkson and Katherine Fowler carried out the process evaluation.

Ethical review

The pattern of headteacher consent followed by parental opt-out consent, as adopted for other EEF literacy catch-up trials run at NFER, was approved by NFER's Code of Practice Committee on 23 January 2013.

Trial registration

This trial was registered at <http://controlled-trials.com/ISRCTN55058116/speaking+listening>.

Methodology

Trial Design

The evaluation was run as a randomised controlled trial involving 235 Year 7 pupils² across three secondary schools who were randomised at pupil level to two groups—intervention and a waitlist control—within the same school. Schools recruited had agreed to allocate children to classes using the randomisation result and then timetable appropriately. For timetabling reasons, pupils were randomised across timetable halves in two schools. Pupils in the intervention group in each school were divided into smaller groups to receive the intervention, comprising the VEIP and the NIP, in withdrawal or timetabled sessions. Pupils in the control group experienced their usual lessons. Pupils were tested for reading ability both before the intervention (after randomisation) and after the intervention, and tested for sentence recall ability after the intervention only. This design sought to determine whether the VEIP and the NIP in combination improved reading ability and/or speaking and listening ability when delivered as an addition to the pupils' usual English curriculum.

When not delivering the intervention, TAs worked with pupils in year groups other than Year 7 to avoid contamination.

Eligibility

The developer, [name], was responsible for recruiting schools, all of which were located in the West London borough of Ealing. This included [his/her] own school, where [he/she] works as an assistant headteacher, and two other local schools.

Eligibility criteria, as stated in the protocol, was defined as: Year 7 pupils that were below National Curriculum level 4 in English and/or below level 4 in reading at the end of Key Stage 2, or pupils that are deemed to be 'vulnerable' Level 4 English achievers, as indicated by either SATs results or a test administered in July 2013.

Given the practical requirement of 80 pupils per school, it was easier for schools simply to identify the lowest 80 attainers in reading using Hodder's digital Access Reading Test.³ The test was delivered by two schools to the new Year 7 intake during the summer holidays of 2013. In the other school, technical difficulties delayed delivering the test until September. It was necessary for timetabling reasons to randomise exactly 80 pupils per school and, in practice, this meant that the eligibility threshold was slightly different for each school. An absolute threshold for eligibility would have resulted in slightly different numbers per school.

Consent

Consent was sought from headteachers who had to complete a Memorandum of Understanding (see Appendix B). Opt-out consent was sought from parents of pupils who met the eligibility criteria and had been selected for participation (see Appendix C). Only three pupils were excluded by parents as a result of the consent letter. These were all from one school and exclusions happened before randomisation.

Intervention

The speaking and listening intervention trialled in this evaluation consisted of sequential delivery of the Vocabulary Enrichment (VEIP) and the Narrative Intervention (NIP) programmes created by

² Although lower than the 240 quoted in the protocol power calculations, using the actual correlation between the baseline and follow-up scores this gives an MDES of 0.29 at 80% power.

³ This test is used annually by the schools and was not used by the evaluators in their analysis.

Victoria Joffe. Two TAs in each of the three participating schools delivered the intervention to their assigned groups of between three and eight Year 7 children (in one school one deliverer was a SEN teacher). All pupils assigned to the intervention group were intended to receive the 24 VEIP sessions in the Autumn term 2013 (beginning in September for 12 weeks) and the 21 NIP sessions in the spring term 2014 (beginning in January for 11 weeks). Each group had two sessions of the intervention per week, scheduled to last between 45 and 60 minutes depending on the school timetable. In one school, lessons were formally timetabled throughout the intervention; at a second, pupils were withdrawn from a variety of other lessons (excluding core subjects such as English, mathematics and science); and in the third, lessons were withdrawn during the VEIP and timetabled during the NIP. The control group attended their usual lessons during the trial period, and are due to receive the intervention next year when they will be in Year 8.

Remaining faithful to how the interventions were originally specified, sessions were designed with a consistent structure, including: a review of the previous session (where applicable) and the pupils' homework (termed their 'Mission to Achieve'); an outline of the current session; the main content; a summary of the session; and setting the pupils' 'Mission to Achieve' for the next session. The training emphasised that sessions should focus on 'facilitation and elicitation' of responses. During training sessions, TAs were encouraged to make the sessions enjoyable to attend by using plenty of interactive activities and to offer positive reinforcement.

Each TA received a folder for each programme which contained the session plans and objectives, additional teaching notes, and session materials (such as template word maps and picture prompts), some of which are included on an accompanying DVD so that they can be used on a projector or interactive whiteboard. Schools were required to provide dictionaries, thesauruses, a microphone (not necessarily a real or functioning one), personal folders for pupils and various everyday objects as discussion stimuli.

Outcomes

The speaking and listening intervention aimed to develop pupils' abilities in recalling vocabulary and in expressing and understanding narrative, on the basis that these language competencies would enhance pupils' attainment in reading. Reading ability was chosen as the primary outcome, as improving reading ability was the main aim of the intervention. The test used was the paper version of the New Group Reading Test (NGRT, GL Assessment),⁴ as it is a reliable test that has been standardised for the age group participating in this evaluation. The NGRT has two forms which have different content but are at the same level of difficulty and provide equivalent standardised scores: Form A was used at baseline and Form B at follow-up. The NGRT has two subscales, sentence completion and passage comprehension, which can be combined into a composite reading score. The composite reading score was used as the primary outcome. The two subscales were used as secondary outcomes.

In order to minimise bias, the project developer, who is based at one of the participating schools, was not involved in testing. In one school, externally recruited invigilators were used for test administration. Otherwise, school staff invigilated while pupils took the NGRT tests under exam conditions. Completed scripts were sent to GL Assessment for blind marking. Results were calculated using GL Assessment's online system, and were accessed through GL Assessment's online platform following testing.

Although the developer was not involved in data collection, complete blinding at NGRT test administration cannot be assured as this was sometimes the responsibility of school staff who were likely to have been aware of the intervention taking place and may have been aware of the allocation

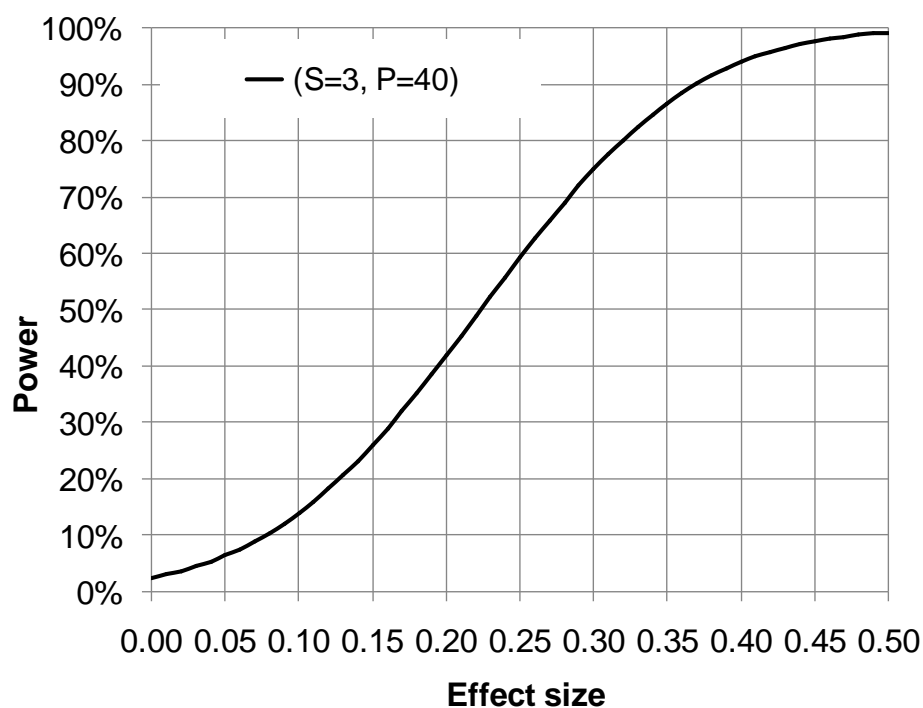
⁴ <http://www.gl-assessment.co.uk/products/new-group-reading-test>

of pupils to the intervention group. Schools were asked to ensure that different staff administered the NGRT test to those that delivered the intervention. It was made clear to the schools that this was a requirement. No formal checks on independent administration were carried out, but five of the six TAs who were available for interview said that they had not been involved in testing.

A Clinical Evaluation of Language Fundamentals (CELF), Pearson Clinical Recalling Sentences Test⁵ was used as an additional secondary outcome. This secondary outcome was chosen as a measure of pupils' language abilities as the intervention aimed to develop pupils' language skills, specifically in speaking and listening (and in turn reading skills, tested by the NGRT as described above). The test used required pupils to work one-to-one with a test administrator and recall sentences. Test administrators were independent and subcontracted by NFER. They recorded and scored the accuracy of the pupils' recall. Administrators were not informed of the allocation of pupils to treatment groups, and were asked not to engage in conversation with staff at the school about this. The results were collated by the evaluator (the data was captured from the scoring sheets).

Sample size

Figure 1: Power curve



S = Number of participating schools

P = Number of participating pupils per intervention group per school

Randomisation was conducted at the pupil level, and variation in baseline scores was controlled for in the final analysis. Intra-class correlation (ρ) was therefore likely to have a minimal impact on the effective sample size, and so we conservatively assumed a value of $\rho = 0.02$ for the purposes of our calculations. We also assumed a correlation of 0.75 between baseline and follow-up scores on the basis of a previous RCT with reading test outcomes (Smith et al, 2007). The power curve in Figure 1 illustrates that a sample size of 240 pupils should be sufficient to detect effect sizes of the order 0.32.

⁵ <http://www.pearsonclinical.co.uk/CELF>

This could be considered moderate, equivalent to around 4 months of progress; quite reasonable for targeted interventions providing support to small groups of pupils.

Minimum Detectable Effect Size (MDES)

Once all the data from the trial was available, the assumed parameters from the above calculations were compared to the actual parameters and included in a calculation of MDES.

Randomisation was carried out at the pupil level thus cancelling out the effect of clustering when estimating internally valid uncertainty around the effect. Rho can hence be regarded as zero. A value of rho greater than zero was assumed in the sample size calculations due to the possibility of addressing external validity but this turned out not to be appropriate. The adjusted R-squared for the primary outcome model without the intervention term was 0.37, implying a value of 0.61 would have been more appropriate for the correlation between baseline and follow-up scores. Using the actual number randomised, this yields an MDES of 0.29 at 80% power.

Randomisation

The developer and the individual schools involved in the trial were responsible for pupil recruitment. Randomisation was carried out by a statistician at NFER using a full syntax audit trail within SPSS (see Appendix D). Randomisation was stratified by school: in one school simple randomisation of pupils into two groups of the same size was carried out, and in the remaining two schools randomisation was carried out within each timetable half. This was necessary to aid timetabling of the sessions within schools.

Schools were sent the results of the randomisation prior to baseline testing for timetabling reasons. The developer was fully aware that this was a departure from ideal practice due to potential bias that might ensue at pre-test. She negotiated with the member of staff responsible for timetabling in her school to try to avoid it but her request to delay results of the randomisation was not granted. Knowing the importance of an unbiased pre-test administration, the developer was involved in close monitoring of the other two schools and arranged for the use of external invigilators in one school; no other arrangements were made to ensure that teachers administering the pre-test were kept blind.

Analysis

The primary outcome was reading ability as assessed by raw scores from the NGRT. Raw scores were used in preference to age-standardised scores due to potential ceiling or floor effects in the latter. Age of pupil was initially included in the model but was found not to be significant. Subgroup analysis on the primary outcome was carried out on the following groups: gender, school, prior attainment in reading and whether or not a pupil was eligible for free school meals (FSM).⁶ The secondary outcomes were the two NGRT subscales: sentence completion and passage comprehension and the CELF (Clinical Evaluation of Language Fundamentals) test. All outcomes and subgroup analyses were pre-specified at the protocol stage, with the exception of the CELF test.

For the purposes of transparency, an ANOVA on post-test scores alone was carried out before the main regression analysis. Note that this does not represent the definitive analysis as it does not control for baseline test score and is hence vulnerable to any chance or attrition-induced imbalance between groups. Furthermore, it does not benefit from the variance explained by the baseline test and is therefore less sensitive.

The definitive analysis was 'intention to treat', reflecting the reality of how interventions are delivered in practice. It was necessary to take school context into account in the analysis due to the restricted

⁶ The full definitions of the two states that this variable describes are: 1) eligible and registered/claiming 2) not eligible or eligible and not registered/claiming.

nature of the randomisation (Kahan and Morris, 2012). Two dummy variables were included in the regression model to represent school; one school was the default category. The definitive primary outcome analysis regressed post-test raw score on pre-test score, school, randomised group, sex, FSM and age in months. Subgroup analysis was carried out using a separate regression model on FSM pupils and by exploring the interaction between randomised group and pre-test score, school and gender. Secondary outcomes were analysed using raw scores in the relevant domains in place of overall reading scores.

A single-level regression model was employed and this represents a slight deviation from the protocol where it was specified that a multi-level model would be used. Because randomisation was restricted by school, it was necessary to account for school in the model but this was done using dummy variables rather than by using a multi-level model. This method was preferred as a multi-level model would not have estimated variance adequately with only three schools. As this was a pupil-randomised trial, clustering for the intervention effect is not an issue in terms of internal validity.

The main analysis was followed by an 'on-treatment' analysis where data from teacher logs was used to determine the extent of each pupil's involvement with the intervention.

Process evaluation methodology

The process evaluation encompassed the whole trial, from start-up in July 2013 to completion in March 2014. The evaluator collected information from a review of the training materials, observations of the training sessions, observations of intervention sessions *in situ*, telephone interviews with the TAs, and a review of the qualitative parts of the 'teaching assistant logs'. These methods were chosen to ensure the evaluation team covered all the different elements of the intervention.

Detailed schedules for the training observations, session observations and telephone interviews were developed to ensure that data collection was consistent and comprehensive. As more than one researcher was involved in observing training and intervention sessions and undertaking interviews, team meetings were held to share information and plan next steps.

Session observations and teaching assistant logs: The evaluator observed one session of the VEIP at a school (November 2013) and observed two sessions of the NIP at a second school (two different TAs, March 2014). Only one observation of each programme was required by the protocol, but in the second school there was the opportunity to observe two NIP sessions running consecutively without overburdening the school. These sessions were chosen by randomly selecting from the participating schools and negotiating the visit dates and time with the schools. TAs used a 'teaching assistant log' pro forma to record an outline of each intervention session (date, duration, content) throughout the trial, and to note any deviation from their intervention plan in terms of content or, for example, the session being taught by someone else. The data on the teaching assistant logs, supplemented by interview data, was used to calculate the intervention dosage, that is, the amount of time that each group of pupils was exposed to the intervention. Information from the teaching assistant logs also contributed to the assessment of intervention implementation and fidelity.

Telephone interviews: These were conducted toward the end or after the end of each programme (depending on TAs' availability) in December 2013/January 2014 for the VEIP, and May 2014 for the NIP. Evaluation staff interviewed all TAs about the VEIP and five of the six TAs about the NIP.⁷ Interviews took approximately forty minutes to one hour to complete, and researchers produced detailed write-ups after the interview. The interview topics included perceptions of the training, delivering the lessons, the resources required, perceived outcomes, issues of cross-contamination, and other issues such as perceptions of scalability.

⁷ One teaching assistant did not return logs to the evaluator and was not available for a telephone interview about NIP due to maternity leave. We understand intervention delivery did occur during the planned weeks.

Observations of training sessions were conducted on one day in September 2013 (VEIP) and on day in December 2013 (NIP). An initial training day in July 2013—an introduction to speech, language and communication difficulties and to vocabulary learning—was not observed.

Impact evaluation

Timeline

Table 1: Timeline

Month	Activity
June 2013	Recruitment and consent from schools.
July 2013	Eligible pupils identified to the evaluator by two of the schools (the third school identified their pupils by 3 September).*
July 2013	Random allocation of pupils in first two schools.
September 2013	Opt-outs from parental consent (6 September; only in school whose eligibility test was delayed; no opt-outs from the other schools).
September 2013	Random allocation of pupils in school whose eligibility test was delayed (9 September).
September 2013	Pre-testing (9–13 September).
September 2013–March 2014	Delivery of intervention programmes.
March–April 2014	Post-testing (31 March–4 April).
April 2014	Additional test (CELF) (14 th –18 th April in two schools; 14 th –18 th April and 28 April–2 May in one school).**

*The third school was not able to put forward their eligible pupils to the evaluator in July due to a technical fault with their digital Access Reading test. The evaluator received these details on 3 September 2013.

**In the third school, tests were administered to approximately half of the pupils during the week commencing 14 April as the other pupils were absent; these pupils were tested during the week commencing 28 April.

Participants

Schools were recruited by the developer during June 2013. As they were local schools, the developer already had links with them. The three schools involved in the project were in the London Borough of Ealing, two were comprehensive schools and one was an academy. Seven hundred pupils joined Year 7 in the three schools in the school year 2012/2013. Of these, 236 pupils were deemed eligible for the study and were consented to be randomised to the intervention or control groups. There was one duplicate case that was not spotted before randomisation. To avoid biased allocation, this pupil was removed from the analysis, so effectively the number randomised was 235.

Table 2: Ofsted ratings for schools

Ofsted rating of overall effectiveness of the school	Number of schools
Outstanding	2
Good	1

Table 3: School type

School type	Number of schools
Comprehensive to 18	2
Academy	1

Table 4: Pupils eligible for FSM

Pupils eligible for FSM	Number of schools
Highest quintile	2
Second highest quintile	1

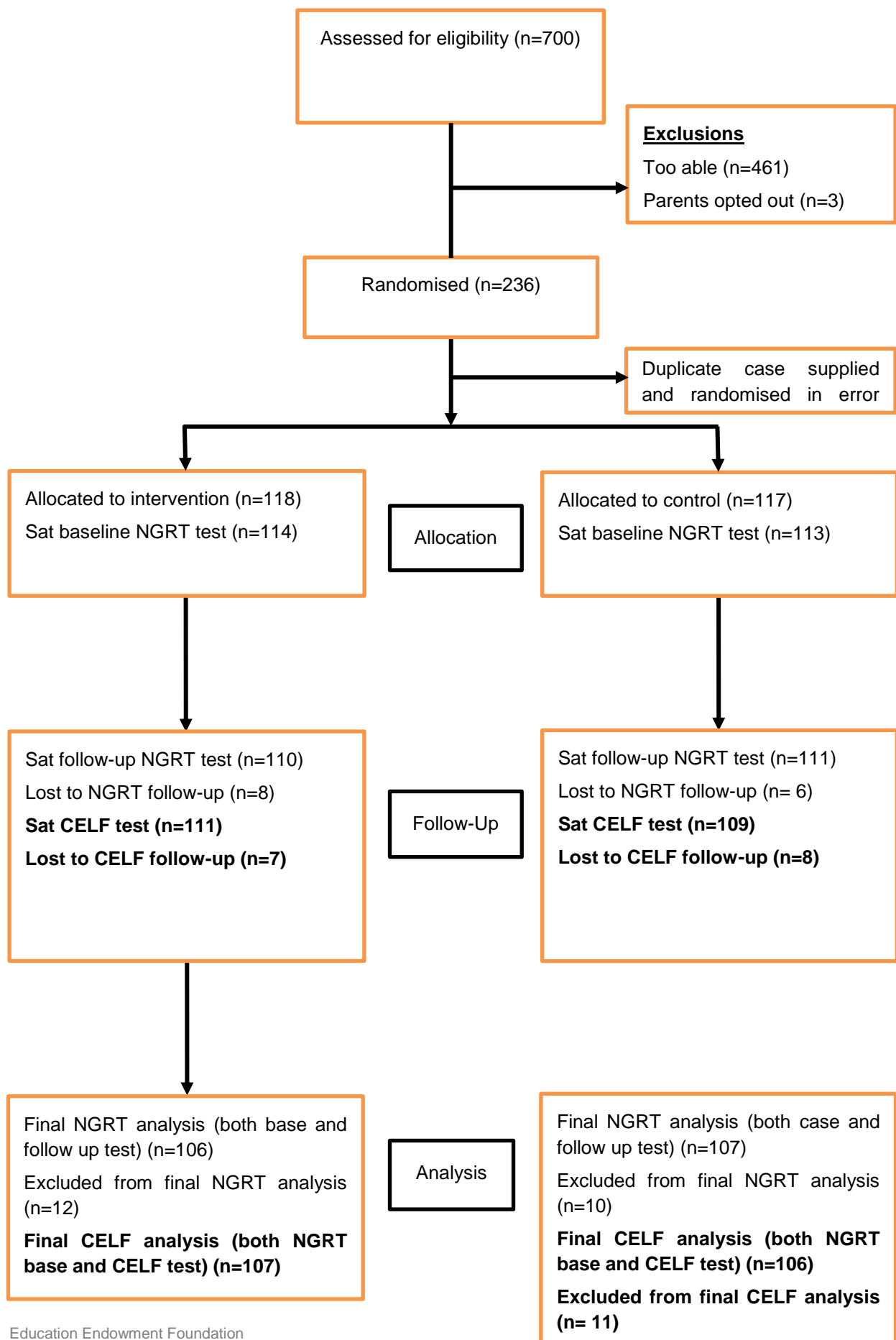
Table 5: School attainment⁸

GCSE performance band (2012)	Number of schools
Second lowest quintile	1
Second highest quintile	1
Highest quintile	1

Two of the schools had an *Outstanding* Ofsted rating and one was rated as *Good*. Two of the schools were in the highest quintile for FSM eligibility and one was in the second highest. Based on their GCSE performance band, one school was in the highest quintile, one in the second highest and one in the second lowest.

⁸ Total GCSE points score averaged for school then weighted by number of pupils in school to establish quintiles. This was used in preference to percentage of pupils with 5 or more A*–C grades as it gives a better overall picture of attainment.

Figure 2: Participant flow diagram



Pupil characteristics of analysed groups

Whilst we expect no systematic bias to have arisen from randomisation, bias may have occurred due to attrition. Chi-squared tests on all three background factors presented in this section revealed no significant differences between groups for the data after attrition.

Table 6: National Curriculum level in reading at baseline ($\chi^2 = 2.47$, $df = 2$, $p = 0.30$)

National Curriculum level (source: baseline NGRT)	Intervention group		Control group	
	Frequency	Percentage	Frequency	Percentage
3	20	19	20	19
4	62	58	63	59
5 or above	16	15	8	7
Missing	8	8	16	15
Total	106	100	107	100

Table 6 shows that the majority of pupils satisfied the eligibility criteria as applied by schools when assessed at baseline.

Table 7: FSM eligibility ($\chi^2 = 0.002$, $df = 1$, $p = 0.97$)

Pupil eligible for FSM (source: NPD)	Intervention group		Control group	
	Frequency	Percentage	Frequency	Percentage
Yes	30	28	30	28
No	76	72	75	70
Missing	0	0	2	2
Total	106	100	107	100

Table 8: Sex ($\chi^2 = 1.70$, $df = 1$, $p = 0.19$)

Sex of pupil (source: schools via GL Assessment)	Intervention group		Control group	
	Frequency	Percentage	Frequency	Percentage
Male	49	46	59	55
Female	57	54	48	45
Missing	0	0	0	0
Total	106	100	107	100

Further to pupil background measures, it was also important to test whether significant imbalance at pre-test had ensued as a result of attrition. The baseline effect size was 0.25 (-0.02, 0.52) and was not significant at the 5% level ($p = 0.064$). Although such a test is not conclusive, this imbalance was probably by chance rather than due to biased attrition. Though not statistically significant, this imbalance is reasonably large and favours the intervention group (see Table 9). Using simple

randomisation, we should expect this kind of imbalance from time to time; it is fully addressed in the regression model by including baseline as a covariate.

Table 9: Pre-test means

Group	Intervention	Control
Mean	23.4	21.6
Standard deviation	7.2	6.7
n	106	107

Outcomes and analysis

Table 10: Effect size

Outcome description	Outcome measure	Effect size (Hedges' g)	95% confidence interval (lower)	95% confidence interval (upper)	p	Number of intervention pupils in model	Number of control pupils in model
Primary	Reading score (NGRT)	0.20	-0.02	0.42	0.07	106	107
Primary (FSM)	Reading score (NGRT)	0.33	-0.06	0.71	0.09	30	30
Secondary	Sentence completion score	0.06	-0.17	0.29	0.61	106	107
Secondary	Passage comprehension score	0.25	0.02	0.48	0.03	106	107
Secondary	Recalling sentences (CELF)	-0.04	-0.28	0.20	0.74	107	104
Secondary (FSM)	Recalling sentences (CELF)	0.16	-0.32	0.64	0.50	30	29

Table 11: Raw outcome means

Outcome description	Outcome measure	Intervention			Control		
		Mean	SD	n	Mean	SD	n
Primary	Reading (NGRT) score	26.0	7.2	110	23.6	6.9	111
Primary (FSM)	Reading (NGRT) score	26.3	6.2	30	23.2	7.3	31
Secondary	Sentence completion score	9.4	2.6	110	9.0	2.4	111
Secondary	Passage comprehension score	16.6	5.3	110	14.6	5.2	111
Secondary	Recalling sentences (CELF)	59.9	12.6	111	59.4	12.0	109
Secondary (FSM)	Recalling sentences (CELF)	64.4	10.9	30	60.7	13.7	30

An ANOVA of the NGRT post-test reading score by randomised group did show a significant impact of the intervention ($F = 6.41$, $p = 0.01$, $n = 221$) but the CELF score did not show a significant impact ($F = 0.068$, $p = 0.794$, $n = 220$).

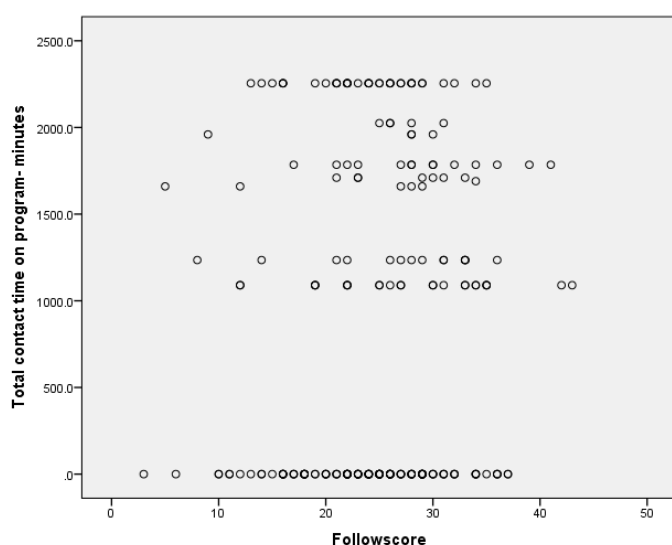
All outcomes analysed were pre-specified in the protocol with the exception of the CELF test. All sub-group analyses were pre-specified in the protocol aside from the use of FSM as a proxy for pupil premium; a separate FSM analysis is a requirement of all EEF evaluations. Background data on pupils was obtained both from schools through the standard GL Assessment data form and from the National Pupil Database (NPD). The main primary outcome analysis consisted of a regression model containing the following variables: pre-test score, intervention group and school. FSM, age in months and gender were included in an initial run of the model but none was significant so all were excluded. Model results are presented in the Appendix.

It is notable that the ANOVA on post-test reading score was significant but the analysis controlling for baseline was not. This has come about due to the imbalance at baseline that favoured the intervention group. Standard deviations for the two tests were similar so raw score differences can be compared approximately. Pupils in the intervention group started with, on average, a 1.8 raw score point advantage (Table 9). At follow-up this had increased to 2.4 raw score points (Table 11). This increase was not enough to cross the 5% threshold of statistical significance when controlling for the baseline imbalance.

No significant effects of the intervention on the primary outcome as measured by the NGRT were found for the following groups: prior attainment ($p = 0.25$), gender ($p = 0.43$), FSM ($p = 0.51$) or school ($p = 0.24$ and $p = 0.34$ for the two dummy interaction variables). In other words the effect of the treatment was the same for pupils of different abilities, for girls and boys and for different schools. The delivery of the intervention varied across schools as it would in a real-life situation, for example, by different teaching styles, lesson duration, group sizes and timetabling arrangements. FSM was explored both using an interaction (quoted here) and as a separate analysis on FSM pupils only (quoted in Table 10) as per current EEF guidelines. Other sub-groups were tested using interaction terms. Of the two secondary outcomes, the intervention was seen to have a significant effect on passage comprehension but not sentence completion. When the CELF test was used as an outcome no significant effects of the intervention were seen.

All the above analysis was ‘intention to treat’. For both elements of the programme, the VEIP and the NIP, the number and duration of sessions was recorded by the TAs who were delivering the intervention. Data was missing for one TA who had gone on maternity leave; rather than exclude these pupils from the analysis they were allocated the same values as the pupils taught by the other TA in that school. The ‘on treatment’ analysis used two measures of intervention experienced by each pupil. The first measure was total contact time in minutes for the VEIP and the NIP combined and varied between 1,090 and 2,255 minutes within the intervention group (see Figure 3). The second measure was the average progress (%) through the course as measured by the number of sessions completed, again for the VEIP and the NIP combined this ranged between 60.7% and 100%. The ‘on treatment’ analysis was carried out by substituting the relevant dosage measure for the dichotomous intervention measure. It revealed no significant effect of dosage for either total contact time ($p = 0.202$) or average progress ($p = 0.146$). See Appendix A for model results.

Figure 3: Scatterplot of dosage versus follow-up NGRT score



Cost

The cost of delivering the intervention to one cohort of 40 pupils each year for two years works out at £29 per pupil. This assumes that a school arranges training for two of its TAs to deliver the intervention, does not need to cover costs for training or for backfilling TAs’ roles, and provides each TA with folders for the VEIP and the NIP. The total cost of hiring a qualified speech and language therapist to deliver the training is approximately £2,000 plus VAT (VAT can be claimed back by schools). Programme folders, which include a resources DVD, cost £87.50 each.

Some costs are not included in the calculation above: for instance schools may wish to buy basic microphones for the NIP, at a cost of about £30 each; photocopying/printing costs are not included; neither are stationery, dictionaries, thesauruses, or everyday objects used as stimulus materials for sessions since these would normally be available in schools. In the trial, TAs chose to use other resources such as tablet computers, voice recorders and word games that were already available in school. Some TAs also chose to buy small rewards for pupils such as pens, stickers and sweets.

Process evaluation

Once the intervention was designed, the next step was to recruit schools and (where applicable) TAs to deliver the intervention: this was completed prior to the commencement of training. The training sessions for the VEIP and for the NIP were delivered at the lead school for the trial by a speech and language therapist from the NHS Trust in Ealing.

The TAs received some introductory training on speech, language and communication difficulties, how pupils might present with such difficulties, and ideas on how to support affected Year 7 pupils (although the trial did not specifically seek to include children with, or exclude children without, these difficulties). This initial training also included vocabulary learning concepts and strategies. This was delivered in July 2013 and was not observed by the evaluator. It aimed to enable TAs to fully access the rest of the training. In addition to the training sessions, TAs met as a group midway through VEIP delivery to share experiences and good practice, although not all TAs were able to attend. They did not meet as a group during NIP delivery. In addition, TAs could contact their own line manager with *ad hoc* queries about the project.

Training observations

Vocabulary Enrichment

The training consisted of a one-day course held in September 2013. All the TAs attended this training. The lead school provided each delegate with a programme folder.

The trainer explored the TAs' expectations of the VEIP training and covered the background and specific aims of the VEIP. Key concepts in the VEIP, including word maps and word pyramids, were discussed in detail. The VEIP includes 24 sessions. The TAs worked in pairs to look through a number of sessions; they then shared and consolidated the learning through group discussion. In the afternoon session, the TAs continued this paired exercise, working through all the sessions and feeding back to the group.

The training gave the TAs a comprehensive overview of the programme and the trainer delivered with clarity. The TAs were given plenty of opportunities to raise questions and were highly engaged in the training.

A basic principle of the VEIP is to gradually build pupils' vocabulary knowledge. The VEIP guidance notes (in the teachers' folder) state that pupils are not expected to move on until they have fully grasped a concept. At the training, TAs voiced concerns regarding how the research trial would be conducted, the quantity of activities that they would have to cover in each lesson and the possibility that they would not have enough preparation time to go over the materials before conducting the lessons. They were unsure whether they would need to cover all the material in each session before moving on, and whether they had to move on regardless of pupils' progress. The trainer did not know all of the requirements of the research trial but followed up these queries with the project developer/manager. The TAs reported that they were subsequently told that to adhere to the time frame of the research trial pupils would need to be moved through the different lessons at pace. The TAs perceived this to be a barrier to delivering the intervention (this is discussed further in the 'Barriers to delivering the intervention' subsection below).

Although the TAs had concerns regarding the practical aspects of delivering the programme, most reported that by the end of the training session they felt prepared and quite confident. One TA did not feel prepared, commenting that the training seemed to presuppose knowledge of speech and language difficulties which she did not have. Another TA, who was very experienced in supporting children with special educational needs, also felt that without pre-existing knowledge of language

difficulties the training would not be sufficient. One TA commented that the introductory session and the VEIP training should have been delivered back to back, instead of having a gap between the two sessions. Not all TAs had received the VEIP folder or had time to read through it before the training.

Narrative Intervention Programme

The training for the NIP took place in December 2013. The training was similar in format to the VEIP training and also lasted one day. One TA could not attend due to illness and planned to review the training materials with her TA colleague before starting delivery. Delegates were mostly very engaged, particularly in the afternoon's activities of session reviewing and planning, although one introductory part of the training was slower-paced and repetitive (compared to the rest of the training) and delegates seems to lose focus.

Three of the five TAs interviewed felt that the training was helpful, if quite discussion-based rather than primarily instructive. As they already had experience of delivering the VEIP, they were confident about delivering the NIP. Two TAs thought that the training in itself was inadequate, that teaching experience (and not TA experience alone) was an important prerequisite to being able to teach the programme effectively. They thought that the following elements should have been included: behaviour management, how to group children according to ability, and modelling of good practice—for example by viewing a video of a TA delivering a session effectively.

Implementation

Each group received two lessons of the intervention per week. These were delivered by TAs (with varying amounts of experience in the TA role) except in one school where one deliverer was a SEN specialist teacher. The lesson duration, group sizes and timetabling arrangements differed among schools. At one school, the scheduled lesson duration was 45 minutes, pupils were withdrawn from other lessons for the VEIP but then had fixed-timetable lessons for the NIP, and group size was six to seven. At a second school, lessons were 50 minutes, pupils were withdrawn from lessons, and group sizes were between five and eight. To avoid an adverse affect on other curriculum subjects, where withdrawals were made, these were arranged to affect multiple subjects (such as humanities, PE, or languages) rather than repeatedly withdrawing the pupil from the same lesson. The third school had 55-minute lessons on a fixed timetable with three to four in a group. Groups made varying amounts of progress through the sessions for each programme, as explained in the Fidelity section. At protocol stage it was envisaged that schools might do things differently, and whether schools were differentially effective was explored in the Outcomes and Analysis section through an interaction between intervention and school.

The observed VEIP lesson was scheduled to last 45 minutes with seven pupils. Most pupils were five minutes late in arriving, two pupils were ten minutes late, and one was absent. The TA said that the pupils needed time to get from one lesson to another and this cut into delivery time. The classroom allowed the group plenty of space and there were no distractions. The lesson included several short activities from the relevant plan in the programme sequence (session 8 out of 24) and the TA moved through them at a pace that kept pupils engaged. The objectives included revising the alphabet, grammatical terms and parts of words/speech, and describing objects focusing on similarities and differences. The lesson was highly interactive: for instance, the TA and pupils threw a ball to each other to nominate someone to give a definition or example; pupils also worked in pairs to build words using blocks; and pupils recorded themselves taking turns to say the alphabet and then played it back. This kept pupils interested and in particular helped to include the minority of very quiet pupils who seemed to find writing tasks particularly challenging (the amount of writing done was minimal). Behaviour was generally very good although there was some off-topic conversation. Concepts that pupils were struggling to retain were recapped more than once during the lesson, and accurate responses were encouraged. Pupils became quicker to recall concepts over the course of the lesson and most were keen to answer questions and offer examples.

The two observed NIP lessons were scheduled to last 50 minutes each. These were session 15 and 16 of the programme of 21 sessions. The first session with six pupils ran on time. The classroom was quite small for pupils to move around when doing pair work or presenting. There was not a clearly defined lesson structure, but activities moved on at pace. Some discussions were cut short, but all pupils were able to participate meaningfully. Pupils reflected on what they had gained from the intervention so far; they mentioned improved confidence in speaking and improved reading ability. The TA led discussions on the importance of self-evaluation in their story writing, and of active listening. Pupils each presented their Mission to Achieve (homework) and most narrated very confidently with good structure and tone. They then planned the key features of a story. Many found it difficult to just list the key features as they were absorbed in the detail of the events and characters in their stories. The TA asked pertinent questions, and referred to learning from previous sessions and explained how they would build their skills in future sessions.

The second observed lesson with seven pupils ran a few minutes late, and two pupils arrived five minutes after the start. The room set-up made it rather awkward for the pupils to get to the front to present, which took up time. The lesson felt quite rushed and 'bitty' as there were several activities to cover. Pupils generally responded well to the pace, however. One activity was introduced and then seemed to be abandoned. Self-evaluation was discussed very briefly. Pupils presented their homework—character word maps and descriptions. They were asked to map out a story in bullet points in small groups, and to think of adjectives and so on to make the outline more interesting. Pupils struggled to do this in the time available for the activity. Pupils' behaviour deteriorated during a task where they were meant to present their mapped-out story; some insisted on acting it out in their groups rather than simply presenting it, but lost focus. The final activity dealt with communicating emotions, and each pupil was asked to say a sentence to the group in a certain tone. Some were clearly uncomfortable about this, and the TA encouraged them but did not insist that they took part.

In all observed sessions the enthusiasm of the TAs and pupils was evident, and creativity and positive feedback were emphasised. In the NIP sessions, TAs did not always insist on high standards of behaviour, check pupils' understanding effectively or give feedback to improve weak responses. One TA commented that pupils' behaviour had worsened toward the end of term when pupils were becoming tired and their interest was beginning to wane.

Facilitators to delivering the intervention

Support to plan and deliver the programme

TAs found it helpful to work with their co-deliverer at their school, to share planning and ideas. They also appreciated support from their Special Educational Needs Coordinator (SENCo) or line manager.

The programme activities

The programme folders had helpful elements, although they were not ideally tailored to the intervention as it was delivered for the trial (see Barriers section for further discussion). One TA commented that the content included 'plenty of fantastic activities'. The activities which best engaged pupils tended to be creative, interactive and/or competitive, such as using spider diagrams to link concepts with examples; word association tasks; quizzes; word games; role play; enacting certain words, and so on. Some topics within the VEIP were particularly popular with pupils—such as nutrition and health, Earth and space, emotions and careers—as these were topics that pupils 'could relate to' providing a balance with some of the more technical learning in the first half of the programme.

The active listening component of the NIP was regarded by the TAs as essential for many pupils who lacked this social skill. Making a poster of the listening skills as a point of reference for pupils helped

to manage behaviour in future sessions. Sessions where pupils evaluated their own and peers' work were also felt to be effective. Pair work, for instance in reviewing creative writing, helped engage pupils who were reluctant to contribute to whole-group discussion, and demonstrated pupils' understanding.

Pace of delivery

It was important to teach at the pupils' pace and include a lot of repetition and consolidation to ensure that pupils could feel they were achieving something. This was challenging, however, because of the amount of content TAs perceived they needed to cover (see 'Barriers to delivering the intervention' for further discussion).

Timetabling and location

Having a fixed timetable, rather than withdrawing pupils from lessons, helped to improve attendance. It was helpful to use classrooms with sufficient space for activities such as role play, particularly in the NIP.

Barriers to delivering the intervention

Planning and preparation time

The programme materials were perceived to be very 'wordy' and lengthy. Distilling the resources into achievable lesson plans took considerable time. Deliverers estimated that preparation for a session (which could be delivered to multiple groups) took between 30 minutes and three hours, depending on the familiarity of the material and resources required. They had to type-up tasks from the programme folders into worksheets as they felt the materials were not ready to use in class.

Limited time to deliver content

TAs agreed that there were many useful activities within the programmes but the amount of content was very challenging. They thought that they needed to get through all of the sessions for the purposes of the trial. They selected from the activities in each session, but as the programme is designed to be delivered sequentially and cumulatively, skipping activities occasionally became problematic further on in the intervention when a session was designed to build on prior learning. There was a lack of time for reflection on the pupils' progress and for feedback to pupils. They reported also that their pupils needed to go at a slower pace, with more repetitions and consolidation than time allowed.

Even those TAs with relatively small group sizes (three or four compared to TAs with seven or eight pupils) struggled for time when pupils individually gave feedback or presented to the group; pupils did not have time to fully express what they wanted to say.

Appropriateness of the content

TAs reported that not everything worked well or seemed appropriate for their pupils. One TA felt strongly that the programmes had too many broad aims and that the material was not specific enough to any one literacy problem. Some activities (such as the learner profile and more open-ended homework tasks) were too complex or unstructured for the lower ability pupils, and others were too basic for the higher ability pupils who began to be bored. The science content in the VEIP, for example, was generally felt to be 'pitched at too high a level' for the pupils.

Timetabling, attendance and punctuality

One school had significant problems in organising the intervention timetable/rooms, resulting in the VEIP starting some weeks later than intended. At two schools, the teaching space available was not ideal: at one school, some sessions took place in the canteen whilst it was not in use; in another, pupils had difficulty accessing classrooms used for the NIP. One TA had to teach the same group twice in one day so the 'Mission to Achieve' (homework) also had to be fitted into the lesson time. At protocol stage it was envisaged that schools might do things differently, and whether schools were differentially effective was explored in the Outcomes and Analysis through an interaction between intervention and school.

Attendance rates tended to be poorer for withdrawal (compared to timetabled) sessions, as children needed to leave other double lessons part way through. Some lacked confidence to ask to leave the class, or were disorganised and forgot to attend. Some groups regularly started sessions late because pupils had to walk across the school site to reach the classroom. Absence, late starts and lessons missed through school-wide activities and strike closures reduced the time available for the intervention (TAs rearranged missed sessions where possible). Attendance was much higher overall for the NIP than for the VEIP, at least partly due to schools' timetabling changes (for example, one school changed to timetabling the sessions rather than withdrawing pupils).

Pupil groups

Most TAs thought groups were too large for pupils to get the most benefit from the intervention, due to the lack of time for feedback. Two deliverers were particularly concerned about the range of needs across pupils in each of their groups. They felt some pupils were higher ability than their baseline test score suggested (for instance, because they had not taken the testing seriously), were too dominant in the group and at times were becoming bored through lack of challenge.

Behaviour management

Observations indicated that behaviour management was an issue during some of the delivery, specifically when pupils did not follow (or misinterpreted) instructions when doing small-group work and did not do the task as intended. The evaluator saw only three sessions of the intervention, however, so this was not necessarily a general problem.

Fidelity

TAs selected from the activities set out in the programme folder to create a manageable lesson. Their lesson plans did not appear to be biased towards particular types of activities or topics, and they ensured that they 'laid the foundations' of the intervention by working through the first ten or so sessions of the VEIP in detail. TAs variously made adaptations to some activities to make them more engaging or appropriate for their groups' abilities. For instance, they used PowerPoint presentations, created quizzes, focused on verbal communication (rather than insisting on pupils writing down answers), or simplified the wording of some activities. One school did a brief survey of pupils to gauge their confidence with the various aims of the NIP, and tailored sessions accordingly. The school also set targets based on the aims which encouraged pupils to take their progress seriously.

All lessons were taught by the intended TAs. If the TA was unavailable, the lesson was missed and (for withdrawal lessons) pupils instead attended their timetabled lesson. It was not always possible to reschedule missed lessons. Due to missed lessons and the slow progress of some groups, different groups reached session numbers ranging from 12 to 24 (the final session) in the VEIP intervention, and from 15 to 21 (the final session) in the NIP.

Outcomes

TAs felt that the time restrictions of the trial and the mixed-ability groupings had prevented pupils from getting the full potential benefit in terms of their speaking and listening skills. TAs did, however, think that there were positive outcomes for pupils in terms of:

- ability to learn and retain vocabulary, and to make connections between words;
- ability to find words in a dictionary;
- communication skills (interviewing, presentation, narrative and active listening);
- confidence, clarity and expression in verbal communication;
- understanding of a narrative structure, time, place and character and hence improved storytelling; and
- creative thinking.

TAs felt that improved verbal communication skills, and confidence, were the main outcomes of the intervention, rather than literacy (reading and writing) skills.

In terms of wider school outcomes, the withdrawal arrangements and missed lessons or assessments apparently caused some inconvenience and annoyance for other teaching staff.

Formative findings

Overall, TAs reported that the intervention is conceptually sound and enjoyable to deliver, but several practical issues need to be overcome to make it more beneficial for individual pupils and not prohibitively time-consuming for TAs.

Training topics

The training should clarify the expected coverage of the content, so TAs do not feel they need to rush through it. Programme materials should be made available to TAs in advance, as they were for the trial, so that they can familiarise themselves with them before training. It is useful to include an introductory session on common speaking and listening difficulties shortly before training on the main programme content. TAs without experience of speech and language difficulties may benefit from more input from the speech and language therapist at intervals during the intervention. TAs who have not previously taught pupil groups on their own would benefit from guidance on behaviour management and how to deliver a lesson effectively.

Grouping pupils according to ability

The programme may be delivered more manageably with small groups (2–4) with similar needs and abilities in literacy. The selection and differentiation of activities can then be better tailored to the pupils involved. For some sessions, one TA used a carousel of activities so that pupils could move through activities at their own pace; this approach may be suitable for motivated pupils.

Planning and preparation time

TAs should be allocated sufficient time to plan and prepare for sessions, taking into account the time it takes to condense the programme sessions into achievable lesson plans.

Intervention duration

If a school wants to deliver the entire intervention, the schedule should allow time for plenty of consolidation over multiple lessons and for formative assessment. Schools might consider having longer lessons, or three lessons in a week instead of two as in the original pilot of the programmes.

Extending the intervention over a longer period of time might increase pupils' 'interest fatigue', as described by one TA. Timetabled, rather than withdrawal, lessons are more likely to promote good attendance and punctuality.

Programme resources and activities

TAs were unanimous that the programme materials were too 'wordy' and not designed in a user-friendly way. It would be preferable to have concise lesson plans and guidance on timings for each activity, as this would significantly reduce preparation time. TAs may need to adapt activities in the programme to the needs of their pupils (for example, by making homework activities more structured for less able learners). TAs would welcome support to enable them to link pupil progress in the intervention to their schools' standard assessment measures.

Control group activity

TAs reported that cross contamination between the control and intervention groups was very unlikely. They did not think that pupils in the intervention group had shared their learning with the control group pupils. Two TAs (two different schools) had shared some of the intervention activities with other staff, in the context of talking about techniques to involve pupils in discussion or by showing other staff resources that they had created for the sessions. They thought that other staff would probably not have used activities specific to the intervention during the trial period as those staff would have prepared their own lesson materials in advance.

School staff reported that intervention and control group pupils had equal access to other in-school interventions, such as Catch-Up Literacy, speech and language therapy, and one-to-one EAL (English as an Additional Language) tuition.

Conclusion

Limitations

Two TAs at two different schools had discussed some intervention resources or techniques for engaging pupils with non-intervention English teaching staff, despite being aware of the requirements of a controlled trial. It is not known whether these other teaching staff used the ideas in their lessons, but there is a risk that contamination of the control condition occurred to a limited extent. TAs thought it was highly unlikely that pupil-pupil contamination across groups would have occurred.

TAs were aware that they should not be involved in administering the tests and reported that they had not been present during testing, but as the evaluators did not observe the testing we cannot be sure that tests were administered blind. Overall attrition between randomisation and analysis was 9% and there was no evidence of attrition bias although this was not explored formally using multiple imputation or sensitivity analysis due to a limited budget. Due to timetabling requirements, randomisation happened before pre-testing which may have led to bias in this administration but this is very unlikely as teachers would have had to explicitly note who was in each group and interfere with the testing. The developer was aware of the importance of an unbiased administration.

There was no indication of a significant relationship between the 'amount' of intervention that pupils received (either in terms of contact time or progress through the programmes' sessions) and their test scores. Some of this dosage data was missing, and some TAs recorded their session timings more precisely than others, so we cannot be sure that there is no relationship between dosage and impact because of the relatively poor quality data. The TAs reported that they felt more time with the pupils would result in greater benefits.

This trial was run as an efficacy trial in three secondary schools in the West London borough of Ealing. As it was delivered by TAs (mainly) in schools rather than a developer with specific knowledge of the intervention, the result is more likely to be applicable to a real-world scenario. However, whilst internally valid, this trial has little external validity from a strict statistical perspective. The schools recruited into the trial were local to the lead school (and therefore not randomly selected) consequently the sample cannot be said to be representative of any population of schools beyond the locality. There is, however, no reason to conclude that similar pupils elsewhere would have reacted differently to the intervention.

Interpretation

The main result of the trial was that the speaking and listening intervention had an effect size of 0.20; roughly equivalent to three months of extra progress in reading. However, the result was not statistically significant at the 5% level ($p = 0.068$). In combination with the significant effect on passage comprehension ($ES = 0.25$, $p = 0.03$), it can reasonably be concluded that the intervention had some kind of effect. The NGRT was chosen to test reading comprehension skills because improved reading was the ultimate aim of the project, hypothetically resulting from improving pupils' speaking and listening skills. TAs observed that improved verbal communication skills, and not literacy/reading skills, were the principal immediate outcomes of the intervention. While TAs reported that pupils' confidence in communication had improved, the CELF Recalling Sentences test results showed no impact of the intervention on sentence recall (measuring the accuracy of their verbal communication).

Pupils in the intervention group did significantly better in passage comprehension (a sub-scale of the overall reading outcome measure) than the control group. Elements of the VEIP that may have contributed to this effect include enriching pupils' vocabulary and teaching techniques to recall words and their meanings, and improving their understanding of figurative language. The NIP is also likely to

have contributed to the effect on passage comprehension as it aims to improve familiarity with the structure and forms of narrative and the linguistic devices used by narrators. The skills taught through the VEIP seem closely aligned to the skills needed in the NGRT sentence completion tasks, yet there was no intervention effect on sentence completion. The VEIP does not include substantial reading or writing activities but teaches mainly through interactive activities; pupils may have found it difficult to apply what they had learned through these programme activities to the paper-based NGRT test.

The pilot of the VEIP and the NIP (Nuffield Foundation, no date) reported an impact on vocabulary and narrative skills respectively. There are several possible reasons why the results of this trial differ from those reported findings: these include the characteristics of the study participants, the delivery schedule, and the testing used. In the pilot, most of the pupils involved had specific language impairment but did not have significant non-verbal language problems. The pilot intervention was therefore more immediately relevant to their particular needs than in this trial (where pupils' common issue was under-attainment in literacy) and their underlying needs may not have been adequately addressed by the emphasis on speaking and listening. It is also possible that, as the VEIP and NIP were designed to be delivered simultaneously rather than sequentially, this may have altered their effectiveness.

In the pilot of the programme TAs delivered three sessions of the intervention per week for six weeks, to groups of two to six pupils with speech, language and communication needs. Sessions lasted 50-60 minutes. Compared to the pilot, therefore, some pupil groups in this trial were larger, pupils' lesson times were shorter (in two schools out of three), and they received lessons less frequently (twice a week) but over a longer interval. The greater intensity of intervention delivery in the pilot may have been beneficial; additionally, the support those deliverers had from speech and language therapists throughout the intervention may have been an important component that was missing in this trial. The outcome measurements used differed: the pilot used six non-standardised tests to measure language ability (plus an IQ test), including the CELF Recalling Sentences subtest used in the trial reported here, but not including the NGRT. It is unclear which tests indicated the impact of the pilot because the data and report have not been published.

The process evaluation highlighted that TAs usually struggled to cover the very substantial content suggested for each session in the time available and therefore chose to select a smaller number of activities from the programme folder in order to maintain a pace of session that the pupils could keep up with. Additionally, pupil absence and lateness often contributed to difficulty in completing all sessions. TAs reported that they lacked time to reflect on pupils' progress and ensure learning was consolidated. For example, in one school pupils only covered half of the VEIP sessions. There was no association between dosage and reading performance, but the selection of activities and pace of the sessions may have limited the potential for impact.

Poor behaviour may have affected some intervention groups' progress. The TAs did not report behaviour as an issue—other than pupils becoming tired and losing focus toward the end of each programme and school term—however, observations suggested that behaviour management was an issue during some of the delivery.

Future research and publications

A better appreciation of which component, VEIP or NIP, leads to any improvement in reading—and indeed specific aspects of communication—would be helpful. This could be investigated with further small scale efficacy trials. Furthermore, it is possible that the number of Level 4 pupils in the sample for this trial meant that recipients were too able to experience real benefit from the intervention. A trial across more schools focusing on Level 3 may be justified. However, the intervention was not working differentially for lower ability pupils (as demonstrated by the non-significant pre-test interaction) so it is debatable whether this would yield larger effects.

References

NFER (2014) Vocabulary Enrichment Intervention Programme: Evaluation report and executive summary. Slough: NFER.

Joffe, V. (2011) 'Secondary school is not too late to support and enhance language and communication'. Afasic. Available at: [http://www.nuffieldfoundation.org/sites/default/files/files/NEW%20Victoria%20Joffe%20Afasic%20NL%20Winter%202011\[1\].pdf](http://www.nuffieldfoundation.org/sites/default/files/files/NEW%20Victoria%20Joffe%20Afasic%20NL%20Winter%202011[1].pdf)

Kahan, B.C. and Morris, T.P. (2012) 'Reporting and analysis of trials using stratified randomisation in leading medical journals: review and reanalysis'. BMJ, 2012: 345. DOI: <http://dx.doi.org/10.1136/bmj.e5840>, available at: <http://www.bmj.com/content/345/bmj.e5840>

Nuffield Foundation (no date) 'Enhancing Language and Communication in Secondary Schools (ELCISS)'. Available at: <http://www.nuffieldfoundation.org/enhancing-language-and-communication-secondary-schools-elciss>

Smith, P., Styles, B. and Morris, M. (2007) 'Evaluation of RAPID'. Report to client.

Acknowledgements

The evaluation team is sincerely grateful to Mia Pye who recruited schools, arranged training for the TAs, assisted with communication with schools regarding all aspects of the trial, kept NFER up-to-date with progress of the trial and provided other material for the report. At NFER, the statistical analysis was undertaken by Sally Bradshaw. Pippa Lord commented on a final draft of the report. We are grateful for the comments and additions provided by Calum Davey and Sarah Tang at EEF, two anonymous referees and Mia Pye.

Appendix A: Model Results

Results of main effect model (NGRT):

Coefficients ^a					
Model		Unstandardized Coefficients		Standardized Coefficients	
		B	Std. Error	Beta	
1	(Constant)	11.080	1.442		7.682
	Basescore	.612	.057	.592	10.817
	intervention	1.444	.787	.100	1.836
	school1	-1.877	.969	-.121	-1.936
	school2	-.670	.943	-.044	-.711

a. Dependent Variable: Followscore

Neither female, FSM nor age in months was significant so these were excluded from the model.

Results of FSM model (NGRT):

Coefficients ^a					
Model		Unstandardized Coefficients		Standardized Coefficients	
		B	Std. Error	Beta	
1	(Constant)	8.554	2.433		3.515
	Basescore	.712	.097	.685	7.378
	intervention	2.299	1.344	.164	1.711
	school1	-2.785	1.609	-.189	-1.731
	school2	-.491	1.595	-.033	-.308

a. Dependent Variable: Followscore

Results of secondary outcome model (NGRT – Sentence completion):

Coefficients^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.858	.610		6.319	.000
	BaseSCscore	.474	.053	.530	9.022	.000
	intervention	.150	.296	.030	.507	.612
	school1	.799	.367	.147	2.176	.031
	school2	.539	.356	.102	1.513	.132

a. Dependent Variable: FollowSCscore

Results of secondary outcome model (NGRT – Passage comprehension):

Coefficients^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	9.007	.972		9.265	.000
	BasePCscore	.531	.063	.485	8.384	.000
	intervention	1.361	.624	.126	2.181	.030
	school1	-2.693	.767	-.231	-3.509	.001
	school2	-1.251	.749	-.111	-1.669	.097
	female	1.225	.622	.113	1.968	.050

a. Dependent Variable: FollowPCscore

Results of main effect model (CELF):

Coefficients^a					
Model		Unstandardized Coefficients		Standardized Coefficients	Sig.
		B	Std. Error	Beta	
1	(Constant)	39.917	2.880		.000
	Basescore	.752	.110	.426	.000
	intervention	-.502	1.531	-.020	.744
	school1	1.561	1.883	.060	.408
	school2	3.501	1.862	.135	.061
	fsm	4.159	1.703	.151	.015

a. Dependent Variable: celfscore

Results of FSM model (CELF):

Coefficients^a					
Model		Unstandardized Coefficients		Standardized Coefficients	Sig.
		B	Std. Error	Beta	
1	(Constant)	41.617	5.948		.000
	Basescore	.820	.225	.444	.001
	intervention	2.058	3.028	.083	.500
	school1	3.025	3.654	.120	.411
	school2	.375	3.870	.014	.923

a. Dependent Variable: celfscore

Results of prior attainment, gender and school interaction model (NGRT):

Coefficients^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	8.250	2.116		3.899	.000
	Basework	.686	.083	.663	8.267	.000
	intervention	6.355	2.967	.442	2.142	.033
	school1	-.801	1.386	-.052	-.578	.564
	school2	.226	1.318	.015	.171	.864
	female	1.313	1.115	.091	1.178	.240
	Basework*Intervention	-.134	.116	-.236	-1.157	.249
	School1*Intervention	-2.292	1.949	-.118	-1.176	.241
	School2*Intervention	-1.796	1.895	-.094	-.948	.344
	Female*Intervention	-1.256	1.603	-.077	-.783	.434

a. Dependent Variable: Followscore

Results of dosage model 1

Coefficients^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	11.522	1.480		7.784	.000
	Basework	.610	.058	.588	10.511	.000
	Total contact time on program- minutes	.001	.000	.074	1.279	.202
	school1	-2.186	1.031	-.141	-2.121	.035
	school2	-.794	.976	-.053	-.813	.417

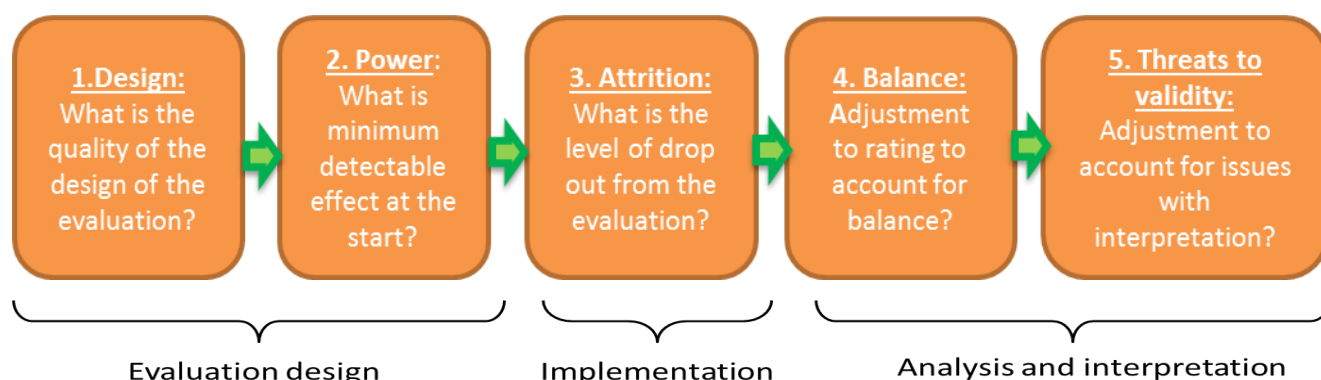
a. Dependent Variable: Followscore

Results of dosage model 2

Coefficients ^a					
Model		Unstandardized Coefficients		Standardized Coefficients	Sig.
		B	Std. Error	Beta	
1	(Constant)	11.465	1.481		.000
	Basescore	.608	.058	.585	.000
	Mean % of progress through course	.014	.010	.083	.146
	school1	-2.063	1.004	-.133	.041
	school2	-.815	.975	-.054	.404

a. Dependent Variable: Followscore

Appendix B: Padlock rating



Rating	1. Design	2. Power (MDES)	3. Attrition	4. Balance	5. Threats to validity
5 🛡️	Fair and clear experimental design (RCT)	< 0.2	< 10%	Well-balanced on observables	No threats to validity
4 🛡️	Fair and clear experimental design (RCT, RDD)	< 0.3	< 20%		
3 🛡️	Well-matched comparison (quasi-experiment)	< 0.4	< 30%		
2 🛡️	Matched comparison (quasi-experiment)	< 0.5	< 40%		
1 🛡️	Comparison group with poor or no matching	< 0.6	< 50%	↓	↓
0 🛡️	No comparator	> 0.6	> 50%	Imbalanced on observables	Significant threats

The final security rating for this trial is 4 🛡️. This means that the conclusions have moderate to high security.

The trial was designed as an efficacy trial and could achieve a maximum of 5 🛡️. This was a well conducted trial, with limited attrition (10%). The MDES was 0.29, which resulted in a loss of 1 padlock. There was an imbalance in the pre-test that was substantial (effect size of 0.25). Because this was accounted for in the analysis, and probably arose due to chance, the security of the result is reduced by only 1. Efforts were made to ensure that test administration was blind, but since tests were delivered internally, complete blindness cannot be guaranteed. Therefore, the overall padlock rating is 4 🛡️.

Appendix C: Parental consent letter

*****, 2013.

Dear Parent/Guardian,

My name is *** and I am ***. We recently secured funding from the *Education Endowment Foundation* for a project which will evaluate the educational impact of Speaking and Listening Interventions for Year 7 pupils. This project is running across several schools in Ealing as well as here at ***** and I am writing to you to make you aware of what is involved in the project. Please feel free to contact me by email or phone if you have any concerns. I can be contacted via email on *** or by telephone on ***

I have attached an information sheet which explains in simple terms what is involved. This is a really good opportunity to help improve speaking and listening, a key part of literacy, and we hope that as many pupils as possible will be able to participate, but we also want to offer you the chance to opt out of the project, if you so wish.

Please return the reply slip at the bottom of this letter to **** should you wish to opt out of the project. If we do not hear from you by this date we will assume that you have no objections and your child will be asked to take part.

Yours faithfully

Speaking and Listening Intervention Project

I do not want my child to be asked to participate in this project.

Child's Name: _____

Parent's Name and Signature:

Office Use only:

Received on:

Appendix D: Example SPSS randomisation syntax

* Randomise pupils within each timetable half.

```
get file='I:/eftr/temp1.sav'.
```

```
freq half.
```

```
select if (Half='X').
```

```
set rng=mt, mtindex=8.
```

```
compute random=rv.uniform(0,1).
```

```
exe.
```

```
sort cases by random.
```

```
Freq random.
```

```
compute lineno=$casenum.
```

```
exe.
```

```
if lineno le 19 group=1.
```

```
if lineno gt 19 group=2.
```

```
ADD VALUE LABELS
```

```
Group
```

```
1 'Treatment Group'
```

```
2 'Control Group'.
```

```
freq group.
```

```
save outfile ='I:/eftr/temp2.sav'.
```

```
get file='I:/eftr/temp1.sav'.
```

```
select if (Half='Y').
```

```
set rng=mt, mtindex=20.
```

```
compute random=rv.uniform(0,1).
```

```
exe.
```

```
sort cases by random.
```

```
Freq random.
```

```
compute lineno=$casenum.
```

```
exe.
```

```
if lineno le 21 group=1.
```

```
if lineno gt 21 group=2.
```

```
ADD VALUE LABELS
```

```
Group
```

```
1 'Treatment Group'
```

```
2 'Control Group'.
```

```
freq group.
```

```
save outfile ='I:/eftr/temp3.sav'.
```

```
add files file='I:/eftr/temp2.sav'/file='I:/eftr/temp3.sav'.
```

```
cross half by group.
```


Appendix E: Cost rating

Cost ratings are based on the approximate cost per pupil of implementing the intervention over one year. Cost ratings are awarded using the following criteria.

Cost	Description
£	<i>Very low:</i> less than £80 per pupil per year.
£ £	<i>Low:</i> up to about £170 per pupil per year.
£ £ £	<i>Moderate:</i> up to about £700 per pupil per year.
£ £ £ £	<i>High:</i> up to £1,200 per pupil per year.
£ £ £ £ £	<i>Very high:</i> over £1,200 per pupil per year.

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 v2.0.

To view this licence, visit www.nationalarchives.gov.uk/doc/open-government-licence/version/2 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 www.educationendowmentfoundation.org.uk



Education
Endowment
Foundation

The Education Endowment Foundation
9th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP
www.educationendowmentfoundation.org.uk