EAL in the mainstream classroom

Tracey, Louise; Boehnke, Jan; Elliott, Louise ; Hanley, Pam; Dysart, Erin; Thorley, Kate

Publication date:
2024

English as an Additional Language in the Mainstream Classroom

Evaluation Report

February 2024

Louise Tracey, Jan R. Boehnke, Louise Elliott, Pam Hanley,
Erin Dysart, Kate Thorley, Sarah Ellison

The Education Endowment Foundation is an independent charity dedicated to breaking the link between family income and education achievement. We support schools, nurseries and colleges to improve teaching and learning for 2 – 19-year-olds through better use of evidence.

We do this by:

- **Summarising evidence.** Reviewing the best available evidence on teaching and learning and presenting in an accessible way.
- **Finding new evidence.** Funding independent evaluations of programmes and approaches that aim to raise the attainment of children and young people from socio-economically disadvantaged backgrounds.
- **Putting evidence to use.** Supporting education practitioners, as well as policymakers and other organisations, to use evidence in ways that improve teaching and learning.

We were set-up in 2011 by the Sutton Trust partnership with Impetus with a founding £125m grant from the Department for Education. In 2022, we were re-endowed with an additional £137m, allowing us to continue our work until at least 2032.

The project was co-funded by The Bell Foundation and Unbound Philanthropy, as part of a funding round focusing on children with English as an additional language.

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

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

info@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Contents

About the evaluator

The project was independently evaluated by a team from the University of York, the University of Dundee, and the University of Leeds:

Louise Tracey (formerly at the University of York, now the University of Leeds), Jan R. Boehnke (University of Dundee, formerly and visiting researcher at the University of York), Louise Elliott (University of York), Pam Hanley (University of York), Erin Dysart (University of York), Kate Thorley (University of York), and Sarah Ellison (University of York).

The lead evaluator was Louise Tracey.

Contact details:

Louise Tracey

University of Leeds
School of Education
Leeds, UK
LS2 9JT

Email: l.c.tracey@leeds.ac.uk

Acknowledgements

We would like to thank the Development Team, including Stefanie Sheddon, Roisin Killick, and Laura Lewis-Williams at Challenge Partners and Li Yen French and Manny Vazquez at the Hounslow Language Service and Lampton School. We also give great thanks to all the teachers and schools who took part in this research and to Kathryn Asbury and Carole Torgerson, at the University of York for their additional support. Finally, thank you to Diotima Rapp (the Education Endowment Foundation [EEF] Evaluation Manager), Katie Luxton (EEF Project Manager), Katharina Keck (EEF Head of Evaluation), Toby Whittaker (EEF Programme Manager) among others at the EEF, The Bell Foundation and Unbound Philanthropy for their guidance and support throughout this research.

The SafePod Network (SPN) was used for secure access to datasets needed for the research detailed in this publication. The SPN is funded by the Economic and Social Research Council and is run by the Scottish Centre for Administrative Data Research. The SPN is part of the Administrative Data Research UK programme. For more information visit: www.safepodnetwork.ac.uk. SafePod is a registered trademark of the University of St Andrews.

This publication includes analysis of the National Pupil Database (NPD): https://www.gov.uk/government/collections/national-pupil-database. The Department for Education is responsible for the collation and management of the NPD and is the Data Controller of NPD data. Any inferences or conclusions derived from the NPD in this publication are the responsibility of the University of York and the University of Dundee and not the Department for Education.

This work was produced using statistical data from the Office of National Statistics (ONS). The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets, which may not exactly reproduce ONS aggregates. ONS agrees that the figures and descriptions of results in the attached document may be published. This does not imply ONS' acceptance of the validity of the methods used to obtain these figures, or of any analysis of the results.

Executive summary

The project

'English as an Additional Language (EAL) in the Mainstream Classroom' was a continuing professional development (CPD) programme for teachers across different subject specialisms to support EAL pupils who are integrated into mainstream classrooms. The CPD programme was designed to build expertise in teachers' planning and lesson delivery for EAL pupils in whole-class contexts through the development of teachers' subject-specific and classroom language use. It is based on an understanding that a better use of grammar, core vocabulary, and spoken language by teachers can help EAL pupils' comprehension and attainment.

This programme was trialled with science and history teachers of Year 10 EAL pupils aged 14–16 years (Key Stage 4) in England who were integrated in mainstream classrooms. A total of 223 teachers in 71 schools took part in this two-armed, cluster randomised controlled efficacy trial, which also involved 5,340 pupils, of which 1,505 were EAL pupils. Schools in the trial were randomly allocated to have their teachers participate in the CPD or continue with teaching as usual.

The CPD programme delivery was led by Challenge Partners, alongside Hounslow Language Service, Lampton School, and regional Delivery Centres. Teachers were required to attend training for three days across Spring and Summer Terms in the 2017/2018 academic year, complete intersessional tasks, and access additional support as necessary. The outcomes were measured at the end of the following 2018/2019 academic year, with the primary outcome measured through General Certificate of Secondary Education (GCSE) science results (combined or the three separate science GCSEs), and secondary outcomes through history and English language GCSE.

The programme was evaluated by a team from the University of York, the University of Dundee, and the University of Leeds. In addition to the impact analysis through GCSE results, lesson observations, surveys, and interviews were used to evaluate the experiences of those participating in the programme. The trial was co-funded by the Education Endowment Foundation (EEF), Unbound Philanthropy, and The Bell Foundation as part of a funding round focusing on raising the attainment of EAL pupils.

Table 1: Key conclusions

Conclusions
EAL pupils in the 'EAL in the Mainstream Classroom' intervention schools made the equivalent of one month's additional progress in GCSE science (GCSE combined or the three separate science GCSEs), on average, compared to EAL pupils in the control group. This result has a moderate to high security rating
EAL pupils in the 'EAL in the Mainstream Classroom' intervention schools also made the equivalent of one month's additional progress in GCSE English language, on average, compared to EAL pupils in the control group. EAL pupils in the 'EAL in the Mainstream Classroom' intervention schools made, on average, no additional progress in GCSE history compared to EAL pupils in the control group
EAL pupils who were eligible for free school meals (FSM) made the equivalent of one month's additional progress in science, on average, compared to EAL pupils eligible for FSM in the control group. However, as the number of EAL pupils eligible for FSM was small, these results should be interpreted with caution
73% of surveyed teachers who received the programme reported they were 'strongly' or 'somewhat' confident in the programme's ability to impact positively on pupils' learning. They also reported increased confidence in supporting EAL pupils, resulting in positive perceptions on EAL pupils' engagement and learning outcomes
The delivery and effectiveness of the programme was likely impacted by delivery challenges including declining workshop attendance, non-completion of assigned tasks, limited support from schools, and a lack of perceived relevance from some teachers

EEF security rating

These findings have a moderate to high security rating. This was an efficacy trial, which tested whether the programme worked under developer-led conditions in a number of schools. The trial was a well-designed two-armed randomised controlled efficacy trial. However, the security rating of the trial was lowered as it was not as well-powered as originally intended because the number of schools recruited was lower than expected. Also, 30% of the EAL pupils who started the trial were not included in the final analysis, largely due to missing Key Stage 2 baseline data for these pupils.

Additional findings

EAL pupils in the 'EAL in the Mainstream Classroom' schools made the equivalent of one month's additional progress in GCSE science (combined or triple science), on average, compared to EAL pupils in the control group. This is our best estimate of impact, which has a moderate to high security rating. However, as with any study, there is uncertainty around the result: the probable impact of this programme ranges from one month's less progress to two months' additional progress.

Secondary and subgroup outcomes tentatively lend further support to the positive primary results of the programme, albeit with similar uncertainty around results. Progress equivalent to one month's additional progress was reported for EAL pupils in GCSE English, disadvantaged EAL pupils who qualify for free school meals (FSM), and for EAL pupils who were taught by more than one teacher who had attended CPD. However, no month's additional progress was found in GCSE history for EAL pupils.

Positive effects in attainment outcomes were supported by findings from teacher surveys and interviews immediately following the programme delivery. The majority of teachers surveyed agreed that the programme impacted positively on EAL pupils' learning. They felt that it had changed their pedagogical practice to better understand the needs of EAL pupils, and that they felt confident in their ability to deliver the programme and support EAL pupils in whole-class contexts.

However, surveys and interviews of teachers also found evidence of limitations to the immediate and sustained impact of the programme. First, the programme's effectiveness was likely impacted by only 60% of teachers attending all training sessions, alongside poor completion, and follow-up of intersessional work. Lack of time was cited as a key challenge for teachers, exacerbated by a change in the curriculum and lack of school support to attend training. Second, some teachers continued to not see the relevance of the programme in their subject teaching, and there was evidence of selective uptake of certain components of the training. These factors could have led to the decrease in teacher perceptions of the efficacy and impact of the programme one year after CPD had ended.

This evaluation adds to the research supporting the achievement of EAL pupils, who currently represent approximately one-fifth of the student population in England. This evaluation tentatively suggests that CPD programmes could be an effective way to improve teaching and EAL pupils' outcomes, at a lower cost than specialist support. At the same time, barriers to the effective support of EAL pupils in mainstream classrooms were also made visible through the evaluation, including teacher capacity, school support, and perceived lack of relevance for some teachers.

Cost

The average estimated cost of 'EAL in the Mainstream Classroom' programme was around £4,558 for one school, or £24 per pupil per year averaged over three years. This is assuming there are 64 EAL pupils per school in Year 10 (the average number of EAL pupils in schools in the programme), and that schools cover costs to cover staff attending the programme training over three days.

Impact

Figure 1: Summary of impact on primary outcome

Outcome/ group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	No of pupils	P-value*	EEF cost rating
Science GCSE EAL pupils	0.06 (-0.06, 0.18)	1		1,071	N/A	£ £ £ £ £
Science GCSE EAL and FSM pupils	0.07 (-0.09, 0.22)	1	N/A	435	N/A	£ £ £ £ £

Introduction

The 2020/2021 school census data (DfE, 2021) shows that learners with English as an Additional Language (EAL) form 19.5% of all pupils in the English education system, which equates to over 1.6 million EAL pupils (Strand and Lindorff, 2021). This represents an increase of almost 3% from that reported by Strand *et al.* in 2015. Whilst there is evidence that some EAL pupils do relatively well compared to their non-EAL counterparts, the overall picture is much less clear with disparities across regions, differences between pupils with different language backgrounds, ethnicity, and those from low socioeconomic status (SES) backgrounds (NALDIC, 2018). However, some research suggests that some groups do significantly less well, with a strong relationship existing between stage of proficiency in English, age of arrival into English education systems, and educational attainment; and that there are strong regional differences and differences between EAL pupils from different ethnic groups in attainment (Hollingworth and Mansaray, 2012; Hutchinson, 2018; Strand, 2016). More recently, research has further investigated whether variation in progression times for EAL pupils may be partly due to variation in assessment practice (Strand and Lindorff, 2021).

A review of language and literacy interventions specifically designed for EAL pupils found a lack of intervention studies within the UK context and within secondary schools (Murphy and Unthiah, 2015). Evidence of small to moderate effect sizes generally came from studies in the US. A recent meta-analysis undertaken by Oxley and de Cat (2021) echoed that explicit vocabulary instruction could yield some gains for EAL learners, including older learners (Crosson and Moore, 2017) and those with the lowest baseline scores (e.g. Hwang *et al.*, 2015). Again, these intervention studies were conducted in the US and pupils were grouped by achievement (August *et al.,* 2014) or English proficiency (Crosson and Moore, 2017). Both meta-analyses also found a lack of continuing professional development (CPD) interventions in this area and, where gains for EAL pupils were found, this was generally after two or more years of intervention (e.g. Matuchniak *et al.*, 2014; Maerten-Rivera *et al.*, 2016).

The lack of CPD interventions for EAL pupils is regarded as 'of particular concern in the UK context given the increasing numbers of EAL pupils in UK schools, a significant lack of EAL pedagogy and too much overlap between Special Educational Needs (SEN) provision and the teaching of pupils with EAL' (Murphy and Unthiah, 2015; p. iv). Strand and Lindorff (2021) suggest that there is a strong relationship between proficiency in English and attainment and schools should be able to assess and moderate this correctly so correct interventions can be applied. Thus, consistent high-quality training in how to assess proficiency should be available to support learners to achieve their highest level of proficiency, which in turn would enable them to achieve their academic potential. Interestingly, one of the highest rated CPD interventions reviewed by Murphy and Unthiah (2015) was based on ongoing workshops to support teachers in integrating literacy and language into the curriculum (Lara-Alecio *et al.*, 2012), thus supporting the significance of ongoing CPD on EAL.

This evaluation of the 'EAL in the Mainstream Classroom' programme was designed to help increase the level of knowledge about possible interventions and the evidence base in the UK, focusing on secondary school pupils, CPD, and changing pedagogic practices to take account of EAL pupils' needs in relation to academic language and attainment. The programme had undergone a London-wide pilot in 58 schools. This initial pilot consisted of five training modules aimed at teachers at different levels of expertise (trainees, teaching assistants, mainstream classroom teachers, lead teachers, and trainers) and was developed by a partnership between Challenge Partners, Lampton School, and Hounslow Language Service. All teachers in the pilot evaluation reported increased levels of confidence in supporting EAL pupils and using a wider repertoire of skills in order to do so. Pupils also reported increased levels of confidence speaking in class as a result of the interventions designed by course participants (Sheddon, 2015). However, the pilot evaluation was not independent (i.e. it was conducted by the programme developers) and there is little quantitative evidence in support of this programme so far. This evaluation involved a two-arm, randomised controlled efficacy trial with random allocation at the school level to evaluate this intervention against usual practice, with an impact evaluation and an embedded implementation and process evaluation (IPE).

Intervention

'EAL in the Mainstream Classroom' is a CPD programme for teachers to support EAL pupils within the mainstream classroom environment, with a particular focus on academic language. Underpinned by wider research, it is designed to enhance teachers' language skills and enable them to provide more focused classroom provision for EAL pupils in particular through improving teachers' awareness of language development, their awareness of the language demands

of their own subjects, their understanding of the specific characteristics of their EAL learners, their competence to incorporate this awareness and understanding into their teaching, and their confidence to change practice.[1] Consequently, it intends to reduce the need for specialist teachers and support staff for this cohort. This is particularly important as schools struggle to provide dedicated specialist support for EAL pupils who are not new arrivals (i.e. those rated C or higher in the proficiency in English scales, see below for further details). The CPD aims to address the lack of consistency in teaching for EAL pupils by improving teachers' skill with language, both general and subject-specific, and provide training in how teachers can plan lessons with EAL pupils' language skills in mind, develop specific resources relating to those skills, and differentiate between pupils with different language skills and varying prior experience of education. The training is designed to support classroom teachers' use and understanding of grammar, core academic vocabulary, and spoken language, which according to the programme's theory of change, are key to helping EAL pupils within a whole-class context, and which are also likely to have benefits for children more broadly.

An additional pilot was conducted prior to this evaluation to assess whether or not to proceed to a trial. This was conducted with the schools recruited as Delivery Centres and trained in the programme during the academic year 2016/2017. The pilot evaluation aimed at assessing evidence of promise, feasibility, and readiness for trial. A report on the pilot findings was presented to the funders in April 2017. The main findings, alongside the criteria for progression to main trial are presented in Table 2. The pilot report is presented in Appendix D.

Although recruitment was not as high as anticipated at that point (April 2017), it was still relatively early in the recruitment stage.[2] Consequently, it was determined that the other criteria had been met and the decision to proceed to main trial was made.

Table 2: Findings of the pilot

Area of evaluation	Evidence criteria	Evidence criteria met
Evidence of promise	• 75% of participating teachers improve understanding of the linguistic demands of their subjects • 75% of participating teachers and programme facilitators express confidence in the programme being able to impact positively on student outcomes	• All (100%) survey respondents agreed that the programme has improved their understanding of the linguistic demands of their subjects • All survey respondents were confident that the programme would impact positively on EAL student outcomes, stating that they were 'extremely' or 'somewhat positive' (100%) • The majority were confident that the programme would impact positively on whole-class outcomes (96%)
Feasibility	• Evidence of programme implementation in more than 75% of classrooms • Evidence that it is implemented with high levels of fidelity, i.e. 75% attendance at CPD and 75% completion of tasks	• There was evidence of programme implementation in all observed lessons • 84% of respondents attended all the training sessions provided • 88% reported that they had completed all of the assignments
Readiness for trial	• Evidence that the Delivery Centres have the ability to deliver CPD to trial schools (including necessary materials and resources in place) and recruit delivery schools	• All the facilitators who participated in the survey were positive about their ability to deliver the programme in the main trial to evaluation schools • The number of schools required to proceed to the main trial has not been met although the numbers of recruited schools is steadily increasing

This evaluation assessed a single unified core module, which focused on mainstream classroom teachers. As an efficacy trial this training model aimed to ensure that the programme was delivered at optimum conditions. Focusing the training on the teachers meant that the evaluation would not need to also take into account differing levels of training of, and assistance by, teaching assistants (as occurred in the previous, London-based pilot). The training was delivered through Delivery Centres located in schools especially selected and trained by Challenge Partners for this purpose. The Delivery Centres, with support from Challenge Partners Delegate Programme Handbook, Hounslow Language Service, and Lampton School, provided three-days' training and support in a group setting to mainstream classroom teachers within their local region in a cascade model to allow teachers to embed new practice. This was a key learning development

[1] See, for example: Dare (2010); Wong and Snow (2000); Carroll (2016); Polias (2010); and Schleppegrell (2010). All references are taken directly from Challenge Partners (2017a).
[2] Recruitment 'officially' started in April 2017 although Delivery Centres had been approaching potential schools prior to that point.

from the previous pilot, which included only one day of training, which was deemed insufficient to enable teachers to embed the learning in the classroom. The training was particularly focused on subject-specific academic language (vocabulary and grammar).

'EAL in the Mainstream Classroom' was intended to be taught across subject specialisms and it was recommended by the developers that science and history teachers from intervention schools who were part of the evaluation be trained with teachers from different subject areas/departments (including teachers outside of the evaluation) to ensure the best training conditions and fullest discussion of language use. However, for the main trial it was agreed to have a single subject specialism (science) as the main focus of the evaluation, with a second, different subject specialism (history) in order to assess the impact of 'EAL in the Mainstream Classroom' as a cross-curricular programme.

Further details of the programme as planned for this evaluation are provided in the Template for Intervention Description and Replication (TIDieR) framework (Table 3).

Table 3: Template for Intervention Description and Replication (TIDieR) framework

Aspect of TIDieR	Exemplification relating to the evaluation
Brief name	EAL in the Mainstream Classroom
Why: Rationale, theory, and/or goal of essential elements of the intervention	A CPD programme for teachers to support EAL pupils within the mainstream classroom environment, with a particular focus on improving academic language (vocabulary and grammar). A pilot evaluation of the programme suggested improved teacher confidence and wider repertoire of skills in supporting EAL pupils as well as increased pupil confidence in speaking in class
Who: Recipients of the intervention	Year 10 science and history teachers and their pupils, with a particular focus on EAL pupils receiving a proficiency in English Grade C or D.[3] Pupils also needed to have been in the English education system at the end of Key Stage 2 (due to the baseline measures, see below) for inclusion in the evaluation
What: Physical or informational materials used in the intervention	Resources included delegate handbooks after each workshop, with links and print outs of useful resources to be adapted for use in the classroom, e.g. the Academic Word List, visual organisation of key information, all targeted specifically for science and history teachers, as well as resources needed for intersessional tasks
What: Procedures, activities, and/or processes used in the intervention	Three days of training workshops within regional Delivery Centres over Spring and Summer Terms 2018 (late January/early February, mid/late March, and late June/early July) and intersessional tasks for participants. This involved improving and understanding academic and subject-specific language, understanding the needs of EAL pupils, and developing and adopting strategies for use within the classroom Strategies related to language use e.g. **Dictogloss,**[4] use of academic language, nouns and verbs (nominalisation), ordering of ideas, 'building the field'[5], increased use of text books, and explaining 'command words' used in exam-style questions. For further detail on the content of the workshops see Table 4

[3] The requirement for schools to report in proficiency in English for all pupils whose language was not described as 'English', 'British sign language', or 'Believed to be English' was briefly introduced into the school census between Autumn Term 2016 and Spring Term 2018. Schools were expected to assess pupils on a 5-point scale of reading and writing. Once an EAL pupil had been assessed as fully proficient in English (code 'E'; Fluent applies) it was deemed unnecessary for the school to continue to re-assess the proficiency of the child on an ongoing basis. However, for all proficiency levels below 'Fluent', it was expected that schools would continue to monitor proficiency on an ongoing basis to ensure that adequate levels of EAL support were provided (DfE, 2018).

[4] As described on The **Bell Foundation's website**:

> Dictogloss is a type of supported dictation. The teacher reads a short, curriculum-related text several times and the learners try to produce their own version as close to the original as possible. The ideal dictogloss text is at a language level slightly above that of the learners, but with familiar subject content. It may introduce some new vocabulary or sentence structures (The Bell Foundation, 2023).

[5] 'Building the field' in this context refers to a series of activities designed to build pupils' background knowledge of a subject or topic. Steps include pre-teaching vocabulary, identifying groups of words in similar fields, identifying the gaps in knowledge, and multiple visits to a challenging text to build understanding and proficiency.

	Intersessional tasks focused on activities to embed learning from the workshops. Further details of these tasks can be found in Table 26
Who: Intervention providers/implementers	Delivery Centres recruited, trained, and supported by Challenge Partners with partners from Hounslow Language Service and Lampton School
How: Mode of delivery	Group-based face-to-face training, across academic subjects and schools within each Delivery Centre
Where: Location of the intervention	Training in Delivery Centres with implementation by science and history teachers in their Year 10 classrooms (to continue, where possible, in Year 11)
When and how much: Duration and dosage of the intervention	Three-days' training, with intersessional tasks. Spring/Summer Terms 2018. Implementation of programme in Year 10 classes over same time period with the option to continue with pupils in Year 11 (academic year 2018/2019)
Tailoring: Adaptation of the intervention	Teachers to adapt strategies according to perceptions of own needs and abilities of Year 10 pupils (aided by inter-workshop tasks in identifying those needs)
How well (planned): Strategies to maximise effective implementation	Workshop delivery over a six-month period to build on knowledge and experience. Registers of attendance maintained. Intersessional tasks. Informal support via telephone provided by Delivery Centres and Challenge Partners

As indicated above, the intervention involved three full days of CPD. The aims of the training workshops are given in Table 4.

Table 4: Aims of training workshops

	Event objectives
Workshop 1	1. Understand the nature of EAL pupils in English schools, the diversity of this subgroup, and how schools and teachers support EAL pupils 2. Understand how to assess the language skills of EAL learners using the revised EAL proficiency code[6] 3. Examine and understand key features of the academic register required by proficient learners of English to attain higher grades at GCSE 4. Apply this understanding to develop strategies for mainstream classes that work for EAL pupils (and others)—focusing on: • Building the field to enable access to complex texts • Understanding how language is built—nouns, noun phrases, noun clauses, and their functions[7]
Workshop 2	1. Examine and understand key features of the academic register required by proficient learners of English to attain higher grades at GCSE 2. Apply this understanding to develop strategies for mainstream classes that work for EAL pupils (and others) 3. Focus on understanding verb forms (tenses) and understanding how they build fluency/proficiency 4. Case study of a student over a year shows development in English aided by interventions, which model: • Verb forms (command/imperative to recount narrative) • Building writing through cohesion (conjunctions, sequences) • Verb forms: modals and past tenses to express opinion 5. Strategies to open access to, and understanding of, challenging texts (figurative and idiomatic language) 6. Building the transition from informal to formal language through nominalisation

[6] For instance, the newly introduced requirement for schools to record pupils' proficiency in English, as detailed above.
[7] Further details on strategies i.e. 'building the field', etc. are described in the section 'Intervention' above.

Workshop 3	1. Examine and understand key features of reading and writing required to attain higher grades at GCSE 2. Apply this understanding to develop strategies for mainstream classes that work for EAL pupils (and others) to develop writing and access to complex texts 3. Understand the distinct characteristics of language used for exam papers 4. Develop understanding of how schools and teachers can support EAL pupils

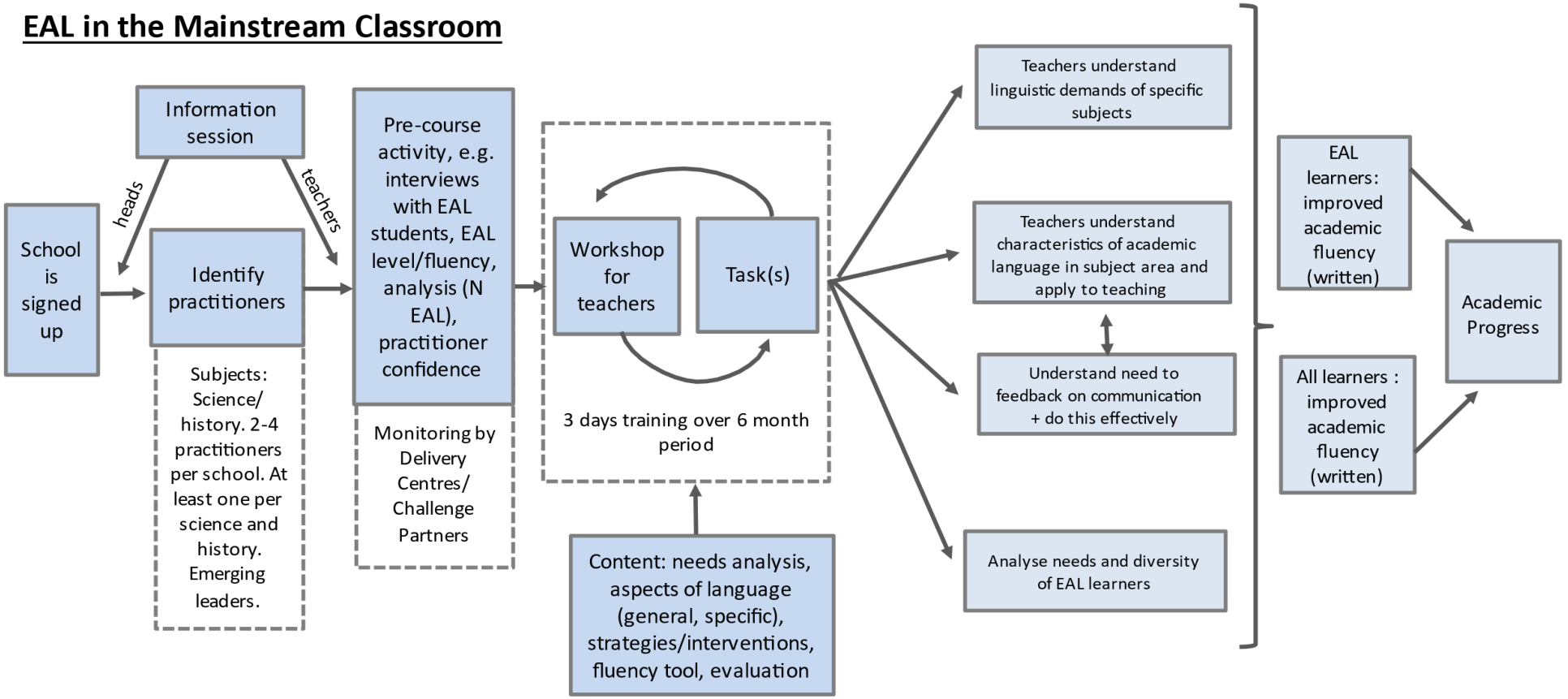
Source: Challenge Partners (2017b).

A logic model for the programme was co-developed by the delivery and evaluation teams and is presented in Figure 2, highlighting the main outcomes measured by the impact evaluation: Improved academic written fluency by EAL learners, and, improved academic written fluency for all learners within the classroom as measured by GCSE examinations. The key inputs influencing these outcomes were identified as:

- increased teacher understanding of the linguistic demands of their specific subject;
- increased teachers understanding of the characteristics of academic language in their subject area and applying that understanding to their teaching/lesson content;
- an increased understanding by teachers of the need to feedback on communication to their pupils and how to do this effectively; and
- increased ability of teachers to analyse the needs and diversity of their EAL learners.

These changes were to be facilitated by the ‘EAL in the Mainstream Classroom’ training alongside intersessional tasks over the Spring and Summer Terms, 2018 to allow for the strategies learned to be embedded into classroom practice though a process of development and deepened understanding.

Figure 2: Logic model



Evaluation objectives

The primary research question the impact evaluation was designed to answer was:

1) How effective is the 'EAL in the Mainstream Classroom' programme in improving subject-specific academic attainment when delivered to Key Stage 4 EAL pupils taking GCSE science?

The secondary research questions, relating to EAL pupils, were:

2) How effective is the 'EAL in the Mainstream Classroom' programme in improving subject-specific academic attainment at Key Stage 4 in a second GCSE subject (history)?

3) How effective is the 'EAL in the Mainstream Classroom' programme in improving academic attainment in English (as measured by GCSE English language) when delivered to Key Stage 4 EAL pupils?

4) What is the impact of 'EAL in the Mainstream Classroom' when pupils receive the approach from more than one teacher in more than one subject area (i.e. when pupils are taught by trained 'EAL in the Mainstream Classroom' teachers in both science and history GCSE subjects)?

A further secondary research question, relating to First Language English (FLE) pupils[8], using GCSE English language attainment as the outcome measure, was:

5) What is the impact of 'EAL in the Mainstream Classroom' on non-EAL pupils within the same classrooms?

Subgroup analyses were also conducted to assess the impact of the programme on EAL pupils with differing baseline fluency bands and on EAL pupils eligible for free school meals (FSM).

The IPE sought to answer the following key research questions:

1) To what extent was the programme implemented with fidelity and what modifications were adopted?
2) What is the impact of 'EAL in the Mainstream Classroom' on the classroom experience?

The logic model (Figure 2 above) was also used to frame the IPE in terms of assessing the extent to which the casual mechanisms actually occurred in intervention schools.

The Evaluation Protocol and the Statistical Analysis Plan (SAP) can be found here and here, respectively. Alternatively see the Education Endowment Foundation (EEF) website: **EAL in the Mainstream Classroom** (**https://educationendowmentfoundation.org.uk/**), which also includes previous versions of the protocol.

Ethics and trial registration

Ethical approval for this study was initially obtained through the Ethics Committee of the Department of Education, University of York (31/01/2017: Ref: 17/01). Initial consent for participation in the trial was through headteacher opt-in using a Memorandum of Understanding, and teacher consent, to ensure teacher buy-in to the evaluation. Copies of these can be found in Appendix E and Appendix F. The request to use pupil data, including linking to the National Pupil Database (NPD) was made through parental information sheets, distributed by schools, with opt-out consent (see Appendix G). It was made clear in all communications to parents and carers that the intervention would be delivered at a class level, so they were not being given an opportunity to withdraw their child from receiving the 'EAL in the Mainstream Classroom' intervention. Due to the introduction of General Data Protection Regulation (GDPR) legislation during the process of this study ethical approval was re-requested (03/09/2018) and an addendum to the original Memorandum of Understanding sent out to all schools in September 2018 setting out the basis for sharing and processing personal data (Appendix H). In line with this, schools were requested to:

[8] The Department for Education data is coded to identify pupils recorded by their schools as having FLE and those where the first language is other than English, i.e. have EAL (Strand *et al.*, 2015, p. 16). The term FLE and non-EAL (i.e. not classed as having EAL) are used interchangeably throughout this report.

- sign and return the addendum to the Memorandum of Understanding (Appendix H);
- distribute updated information sheets to parents/carers (Appendix I); and
- sign and return a Data Sharing Agreement (Appendix J).

The updated information sheet for parent/carers restated the nature of the intervention and the research, provided a GDPR-compliant set of frequently asked questions (FAQs) and reiterated the steps to follow if parents/carers objected to their child's data being shared and wished to withdraw their child from the evaluation. This withdrawal process replaced any reference in the Memorandum of Understanding to opting-out, which was no longer applicable (see Appendix G).

The trial is registered at http://www.isrctn.com/ (Study ID: **ISRCTN15266150**).

Data protection

The legal basis for processing the data was public interest. All school, teacher, and pupil data was treated with the strictest confidence and used and stored in accordance with all applicable data protection laws including the GDPR (EU) 2016/679 Article 6 and the Data Protection Act 2018 (the Data Protection Legislation). Data sharing agreements were put in place between the University of York and each individual school, which included the details of the types of personal data being shared, the purpose and duration of that sharing, and the responsibilities each party had in relation to that information. The University of York, in the form of a privacy notice/participation sheet and an addendum to the original Memorandum of Understanding, provided information to individuals about the use of their personal data. Participants (schools, teachers, and parents of participating pupils) were also provided with the option of withdrawal from the research and details of the process to do so.

The University's privacy notice/information sheet is compliant with the requirements of the GDPR including a clear statement of the University of York's legal basis for processing their personal data, which for this study was under Article 6 (1)(e) of the GDPR: Processing was necessary for the performance of a task carried out in the public interest. Any special category data was processed under Article 9 (2)(j): Processing is necessary for archiving purposes in the public interest or scientific and historical research purposes or statistical purposes. This is in line with the University's charter, which states learning and knowledge will be advanced through teaching and research.

All electronic data transfer, to and from settings, was via encrypted spreadsheets sent through the University of York's secure Drop-Off service. The trial management system and trial data were held separately on secure University of York servers with access limited to specified members of the evaluation team. Confidentiality was maintained and no one outside the research team has access to the study central database. The central database will be stored by the research team for up to one year after completion of the study and then securely destroyed. After the study has been completed the pseudonymised database will be shared with the EEF's archive manager within the Office for National Statistics (ONS) Secure Research Service. For the purposes of the research, data from this central database was linked to NPD data within the ONS. This database was only accessed by the project statistician (Jan R. Boehnke) who currently has ONS Safe Researcher Approval.

The University of York is the data controller during the evaluation up to the point of data being deleted from all locations by the evaluator and/ or delivery team. The University of York is committed to the principle of data protection by design and default and the minimum amount of data necessary for the project was collected. In addition, we anonymised or pseudonymised data wherever possible. All results are anonymised so that no individual schools, teachers, or pupils will be identifiable in the report or dissemination of any results. Results may also be used in presentations and for teaching purposes. Once the project is completed, the data will be archived and once internal quality checks have been successfully completed by the archive manager, the EEF will become the data controller for the dataset.

Project team

The independent evaluation was conducted by researchers from the University of York, the University of Dundee, and the University of Leeds:

- Professor Louise Tracey (Principal Investigator), School of Education, University of Leeds, formerly Department of Education, University of York;

- Dr Jan R. Boehnke (Co-Investigator), School of Health Sciences, University of Dundee and Visiting Senior Research Fellow, Department of Health Sciences, University of York;
- Louise Elliott (Co-Investigator), York Trials Unit, University of York;
- Dr Pam Hanley (Co-Investigator), York Trials Unit, University of York;
- Dr Erin Dysart, Department of Education, University of York;
- Sarah Ellison, Department of Education, University of York; and
- Kate Thorley, Department of Education, University of York.

Additional administrative support was provided at various times throughout the project by Imogen Fountain, Niamh Robinson, and Madeline Crosswaite.

The project team was responsible for the design, randomisation, data collection, analysis, and reporting of the evaluation.

Delivery team

The 'EAL in the Mainstream Classroom' intervention was developed and led by Challenge Partners in conjunction with Hounslow Language Service and Lampton School:

- Stefani Shedden, Programme Director, Challenge Partners;
- Roisin Killick, Programme Coordinator, Challenge Partners;
- Laura Lewis-Williams, Director of Partnerships and Programmes, Challenge Partners;
- Jacquie Smith, Consultant Content Director, Lampton School;
- Li Yen French, Director, Hounslow Language Service;
- Rehana Ahmed, Managing Director, Hounslow Language Service;
- Manny Vazquez, EAL Specialist, Hounslow Language Service; and
- Andy Harvey, EAL Specialist, Hounslow Language Service.

Challenge Partners were responsible for project management, recruitment, quality assurance, monitoring, and administration of the intervention. Hounslow Language Service and Lampton School were responsible for the training design, content development, training delivery, quality assurance, and providing support services.

Methods

Trial design

This was a two-armed school-level randomised efficacy trial. Randomisation between, rather than within, schools was preferred because it minimised the risk of diffusion, which was considered to be quite high, given that the programme focused on an approach to lesson planning and teaching with EAL pupils in mind. The secondary research question relating to the impact of the programme on a second subject area also meant that if a within-school design was adopted, children could be allocated to different conditions across subjects. Intervention schools received the 'EAL in the Mainstream Classroom' training and support free of charge. Control schools were expected to continue 'teaching as usual'.

The evaluation focused on Year 10 pupils. This was believed by the programme developers to give the programme sufficient time to be firmly embedded prior to the end of Key Stage 4 assessments. It also allowed the researchers to avoid introducing any additional burden on classes and teachers relating to those classes in the crucial GCSE year although trained teachers could, of course, introduce the approach to their Year 11 (and other) classes if they wished to do so. Schools were recruited to start participating in the evaluation in the academic year 2017/2018 although there was the expectation that if the intervention teacher was teaching the same class the following academic year (when the pupils were in Year 11) they would continue with the strategies and skills learned through the programme.

Although the intervention can be delivered across subject specialisms, for the purposes of the trial a single subject specialism, science, was chosen as the main focus of the evaluation and pupils' science GCSE results was chosen as the primary outcome measure. Science is a core part of the curriculum, and its study is compulsory to the end of Key Stage 4 (i.e. to age 16) (and therefore is more likely to reach a larger number of EAL pupils). It was also hypothesised by the developers to have a range of components that provide different challenges to EAL pupils, who may understand key scientific concepts but struggle to articulate them. In order to assess the impact of the 'EAL in the Mainstream Classroom' intervention as a cross-curricular programme, history was chosen as a secondary subject (with GCSE history results as a secondary outcome) because this subject is high in contextual language and reasoning and since it is an optional subject it is less likely to be set by ability (unlike science). Also, analysis by the National Association for Language Development in the Curriculum (NALDIC)[9] for 2011 indicated that history was one of the subjects where EAL pupils had lower average point scores at GCSE level compared to other GCSE subjects (NALDIC, 2018). GCSE results were chosen because they are of key interest to schools and are recognised as national markers of achievement. Consequently, science and history teachers of Year 10 pupils were recruited to the trial. An additional secondary outcome was English language GCSE results of participating pupils to assess if there were any indirect effects of the programme on pupils' academic attainment in English language. In addition, using primarily nationally collected data minimises the costs and the burden on schools and pupils. As 'EAL in the Mainstream Classroom' is a whole-class intervention, impacts on EAL and non-EAL pupils were assessed, and an additional subgroup analysis was conducted for pupils eligible for FSM.

Control schools received a financial incentive (£1,500) on completion of all data requirements (i.e. after Summer Term 2019), which could then be used to buy the intervention after the end of the trial if they so wished. This avoided potential problems such as ethical issues if the intervention was not shown to be effective and, as the outcome measure was taken at the end of Year 11, prevented an unnecessary time delay involved in a waitlist design if crossover effects were to be avoided. All teachers (control and intervention) received £25 for completing extra administrative tasks required by the evaluation, including completion of teacher surveys and return of the proposed fluency measure. This was administered by the evaluation team at the end of the main trial (i.e. when all requirements had been completed).

A synopsis of the trial design and outcome measures can be seen in Table 5. Further details on the outcome measures can also be found in the next section.

[9] NALDIC is the National Subject Association for EAL. For further details the website can be found **here**.

Table 5: Trial design

Trial design, including number of arms		Two-arm, cluster-randomised controlled efficacy trial
Unit of randomisation		School
Stratification variable (s) (if applicable)		Delivery Centre / region (via minimisation)
Primary outcome	**Variable**	Science attainment (EAL pupils)
	Measure (instrument, scale, source)	Key Stage 4 GCSE science (NPD)*: • Composite score: average across combined science (two GCSEs) (KS4_APCOMBSCI_91[10]); or • Composite score: average across the three science subjects (biology, chemistry, physics) (three GCSEs) (KS4_APBIO_91, KS4_APCHE_91, and KS4_APPHY_91, respectively) Scale: 1–9 (lowest to highest)
Secondary outcome(s)	**Variable(s)**	History attainment (EAL pupils) English language attainment (EAL and FLE pupils)
	Measure(s) (instrument, scale, source)	Key Stage 4 GCSE history (NPD) (KS4_APHIS_91) Key Stage 4 GCSE English language (NPD) (KS4_APENG_91) Scale: 1–9 (lowest to highest)
Baseline for primary outcome	**Variable**	End of Key Stage 2 reading outcomes (EAL pupils)
	Measure (instrument, scale, source)	Key Stage 2 SATs (KS2_READMRK; raw scores) Scale: 0–50 (lowest to highest)
Baseline for secondary outcome(s)	**Variable**	End of Key Stage 2 reading outcomes (EAL and FLE pupils)
	Measure (instrument, scale, source)	Key Stage 2 SATs (KS2_READMRK; raw scores) Scale: 0–50 (lowest to highest)

** Pupils can either follow a combined science GCSE programme resulting in two GCSEs or a 'three separate sciences' GCSE programme, which would lead to separate GCSE examinations in biology, chemistry, and physics and the award of three GCSEs.*

Changes to the trial design

Given lower numbers of schools recruited to the trial than expected the design was changed (**Evaluation Protocol (amended)**; March 2018) to include the recruitment of a second cohort of schools to participate in the evaluation in the academic year (2018/2019). However, insufficient numbers were subsequently recruited to justify this second cohort and the trial reverted to study a single cohort (**Evaluation Protocol (amended) (2)**; October 2018). Other changes were made to the trial design at the same time as the decision to recruit a second cohort and were retained for the remainder of the evaluation, including: 1) a reduction in the number of case study visits to intervention schools (to account for additional visits to be made to the proposed second cohort[11]; 2) the minimum required number of EAL pupils expecting to take science as a GCSE subject was reduced from 14 to 12 per school and the minimum number of EAL pupils expecting to take history as a GCSE subject was reduced from 14 to 8 per school to facilitate recruitment (**Evaluation**

[10] NPD variable names presented in parenthesis throughout.
[11] The reduced number of case study visits remained in place even after the decision to not continue with the second cohort was taken because the Year 10 pupils in the original cohort were then in their examination year (Year 11) and it did not feel justifiable to visit classrooms under these circumstances.

Protocol (amended))[12]; and 3) the number of teachers recruited to the study per subject per school was reduced from two to one providing the number of EAL pupils taught by that teacher in that subject (science/history) was sufficient to meet the recruitment criteria (all **Evaluation Protocol (amended)**; March 2018). In addition, a mediation analysis based on teacher confidence was deleted due to the reduced sample size and difficulties collecting the data (**Evaluation Protocol (amended) (2)**.

Participant selection

Study participants were teachers of science or history and their Year 10 pupils. The Delivery Centres (who were composed of schools who participated in the pilot for the EAL programme and were trained in the intervention) were responsible for the recruitment of schools for the trial. Each Delivery Centre was located in a different geographical area. There were 12 Delivery Centres in total and each was requested to recruit eight to ten schools. Challenge Partners provided Delivery Centres with a list of schools in its area that had high numbers of EAL pupils. The Delivery Centres were then tasked with approaching these schools with the offer and negotiating their eligibility if they were interested, supported by both Challenge Partners and the evaluation team. Delivery Centres were also requested to use existing networks in their local areas. Eligible schools were recruited by the Delivery Centres in Summer/Autumn Terms in 2017.

Recruited schools were required to:

- initially, release at least one teacher in each of the two subject specialisms for the 2017/2018 academic year who would be teaching Year 10 GCSE classes containing at least 14 EAL pupils taking a GCSE science programme and, ideally 14 EAL students taking a GCSE history programme. In the event some schools were recruited with slightly lower numbers of pupils per subject due to incorrect information at the time of randomisation. However, after consultation with the EEF, it was agreed that schools with 12 pupils per subject per school (and who are taught by an evaluation teacher) should be included. This was subsequently formalised in the amended protocol (**Evaluation Protocol (amended)**);
- be located close to a Delivery Centre to facilitate training; and
- not be implementing the programme or intending to acquire the programme until after Summer Term 2019 if allocated to the control condition.

Participating pupils were those Year 10 EAL pupils enrolled in science and history GCSE classes, taught by the designated teachers trained to deliver the programme, together with their non-EAL peers in the same classes. Pupils had to be in the English education system at the end of Key Stage 2 in order to provide a baseline measure (i.e. only pupils for whom Key Stage 2 scores were available were eligible). EAL pupils were those defined as EAL using the Department for Education binary designation (Yes/No).[13] Between 2016 and 2018, additional data was collected for EAL pupils in the school census relating to their level of proficiency in English (on a scale of A–E). The Department for Education EAL binary designation and the proficiency measure was collected from schools at the beginning of the academic year in which participants were in Year 10. In addition, schools provided the unique pupil number (UPN) for Year 10 pupils whose teacher was participating in the study. Parents were provided with information sheets and given the option of withdrawing the use of their child's data from the evaluation. In the event they did so, any records relating to the student were deleted. The evaluation team alongside the Delivery Centres, were responsible for ensuring the eligibility criteria relating to the number of teachers and pupils were met prior to randomisation. The evaluation team contacted each school at the point of recruitment (i.e. after a signed Memorandum of Understanding was received) to ask for teacher and pupil details in order to check these details.[14]

The initial intention was to recruit 100 schools to participate in the trial. However, there were lower than anticipated recruitment numbers (71 schools recruited) and the eligibility criteria was one of the main reasons identified, which proved more difficult than anticipated for those schools who expressed an interest in the programme to meet. In addition,

[12] The reduction in number of EAL pupils per subject occurred prior to the end of recruitment to Cohort 1, when it was determined that recruitment was lower than expected (with the agreement of the EEF). This was retrospectively formalised in the **amended protocol** published in May 2018.
[13] The Department for Education EAL measure is regarded as 'Yes' if 'A pupil is recorded as having English as an additional language if she/he is exposed to a language at home that is known or believed to be other than English' (DfE, 2020).
[14] In a small number of cases where schools were not eligible to take part, the evaluation team worked with schools to facilitate making them eligible, for example by changing the nominated teachers or increasing the number of nominated teachers. It was not felt acceptable to request classes be reorganised to include the number of required EAL pupils. In a few cases, schools near to, but not quite at, the eligibility criteria relating to the number of EAL pupils were accepted into the trial in consultation with the EEF.

in response to the information on the number and location of the participating schools, two Delivery Centres merged to act as one, and one was wound up after the first workshop due to insufficient training numbers as only one school in its area was participating. The school was reassigned to the nearest alternative Delivery Centre. Conversely, there were too many schools in the Greater London area for the Delivery Centres in that area to service properly, so an additional school was recruited to act as a Delivery Centre that had not taken part in the pilot. Special arrangements were made to train designated teachers from this school to deliver the programme.

As detailed above, initially a second cohort was proposed for the following academic year (2018/2019) to boost the recruitment numbers, but it continued to be difficult to find enough schools who could easily meet the criteria, even when the addition of some flexibility was introduced (e.g. lowering the minimum number of EAL history pupils required). The room for flexibility was limited however, given that the eligibility criteria were driven by the need to ensure that the trial was sufficiently powered to provide robust results. Thus, the decision was taken not to proceed with this strategy (26/10/2018). Consequently, 71 schools in total were recruited to the trial and randomised.

Outcome measures

Baseline measures

The baseline measure was the Key Stage 2 Standard Assessment Tests (SATs) results as these are high in contextual validity and are highly correlated with attainment at the end of Key Stage 4 (cf. Benton and Sutch, 2013). These were collected in Summer Term 2014 and raw scores were obtained from the NPD (KS2_READMRK). The scale of the raw scores is 0–50 and these were treated as continuous data in the analysis. It is recognised that the use of Key Stage 2 results would exclude pupils from the trial who entered the English education system after this cohort had taken the Key Stage 2 SATs (i.e. after the end of Year 6). However, this is not the case for the majority of EAL pupils in schools in England and, at the time of the evaluation, those new to the system with low levels of fluency were often provided targeted support outside of the mainstream classroom. The programme was specifically targeted at those EAL pupils with higher levels of fluency (those achieving proficiency in English bands C and D) although it was designed for implementation at the whole-class level (i.e. all pupils within the class would receive the programme), hence data was collected for all pupils for the evaluation. Although it was not specified in the eligibility criteria, pupils were required to have been in the English school system by the end of Year 6 (i.e. to have taken the end of Key Stage 2 SATs, which were used as the baseline measure) and this was checked by the evaluation team at the point of recruitment.

Primary outcome

The primary outcome is the Key Stage 4 GCSE science score. As a compulsory subject in Key Stage 4, pupils will have taken either combined science, which would result in two GCSE awards, or all three separate science subjects (biology, chemistry, and physics), which would result in three GCSE awards. The GCSE examinations took place in summer 2019 and the results were obtained from the NPD (KS4_APCOMBSCI_91 or KS4_APBIO_91, KS4_APCHE_91, and KS4_APPHY_91). Scores for each GCSE range from 1–9. For the analysis of all questions regarding the primary outcome 'science GCSE', the average of the scores provided was used: over two scores for those pupils taking combined science and over three scores for those pupils taking biology, chemistry, and physics. This variable is abbreviated as KS4SCI in this report. Although the EEF advises caution regarding the use of composite scores in analyses, it was agreed that, in this case, the use of composite scores was justifiable. This was due to a number of factors. First, none of the research questions refer to specific sub-scores of the science exam. Second, the programme aims to embed academic language across the different science subjects. And finally, there is difficulty in separating the sciences given the current assessment system. Consequently, it was hypothesised that an intervention effect should be detectable across all science subjects.

Secondary outcomes

Secondary outcome measures are: 1) the Key Stage 4 GCSE history score for those Year 10 EAL pupils taught GCSE history by participating teachers (i.e. the secondary subject specialism; KS4_APHIS_91); and 2) the Key Stage 4 GCSE English language score for all pupils (this being a compulsory subject; KS4_APENG_91). These two outcomes were chosen to assess the potential impact on a second intervention subject (i.e. GCSE history) and in another subject where the intervention was not delivered, but because of the focus on academic language would be expected to have an impact (GCSE English language). Both measures are scored from 1–9 and were obtained from the NPD.

Additional measures

Whilst the focus of the primary and secondary analysis was on EAL pupils, analysis was also conducted on non-EAL pupils in the same classes. Consequently, information on EAL status was collected from schools by the University of York team at baseline for inclusion in the analysis. The team also collected the information on the pupils' proficiency in English using the EAL proficiency in English bands from the school census from schools at the same time. Between 2016 and 2018 all schools were required to return this information, recording level of proficiency in English (bands A–E, A being New to English and E being proficient), annually for all EAL pupils, so the information would have been relatively up to date (i.e. less than a year old).[15] This data was used for a subgroup analysis on whether or not there is a differential impact of the intervention on EAL pupils in different proficiency in English bands at the start of Year 10. As the proficiency in English bands were determined by the schools themselves, we received this data directly from all participating schools.

Finally, subgroup analysis was conducted for EAL pupils eligible for FSM. Eligibility for FSM was obtained from the NPD. The measure chosen was: EVERFSM_6_P_SPR19, which reflected whether or not pupils included in the evaluation had been eligible for FSM at any point within the previous six years (up to and including 2019). Gender (KS4_Gender) was also obtained from the NPD.

Sample size

The statistical power of the proposed analyses at the original design stage was estimated based on figures achieved in previous, similar trials, using Optimal Design software (Raudenbush *et al.*, 2011). Statistical assumptions were as follows:

- pre-test variance between schools (R^2) = 0.65; and
- intraclass correlation coefficient (ICC) (rho) = 0.19.

These figures were based on figures achieved in previous, similar trials (e.g. Hanley *et al.*, 2016). To reach a minimal detectable effect size (MDES) of 0.20 due to the intervention at a statistical power of 0.80, it was estimated that the trial would need a minimum of 100 schools with a minimum of 14 Year 10 EAL pupils in each subject specialism (i.e. taking a GCSE in history and/or science). Stakeholders agreed that these numbers seemed realistic, since EAL pupils comprised only about 16% of the population at the time of the evaluation design (Strand, 2016). Recruited schools would need to have at least 107 pupils in Year 10 to cover for approximately 20% student dropout during the study. Assuming 15% school dropout (in line with the average for the EEF trials) it was recommended recruiting 115 schools with 17 Year 10 EAL pupils per school. Both were felt likely to be conservative assumptions, although no empirical estimates were available. At the protocol stage, it was noted that this was an optimistic estimate, since recruitment already showed that, beyond this minimal number required for recruitment, the number of EAL pupils within schools varied quite strongly, and the design would have no room to compensate for further dropout. However, given that this was an efficacy trial, and that the developers and Delivery Centres had limited delivery capacity, it was decided that 100 schools would be recruited for the main trial, consisting of schools willing to release at least two teachers in each of the two subject specialisms who would be teaching Year 10 GCSE classes containing at least 14 EAL pupils expected to enrol on a GCSE history programme and 14 EAL pupils taking a GCSE science programme.[16]

In version 2 of the protocol, when the recruitment of a second cohort of schools was proposed it was agreed that a sample size of N = 120 schools, with an average of 14 Year 10 EAL pupils (about 16% of the population; Strand, 2016) in the primary subject specialism (GCSE in science) would potentially detect an effect size of MDES = .22 (significance level $p < .05$, statistical power of 0.80, two-sided test; calculated with formula presented in EEF, 2013). However, given under-recruitment and the decision to not recruit a second cohort for this study the realised sample size for the study

[15] The measure of proficiency was based on teacher judgement, with no external guidance on how to assess pupils' proficiency. Any use of this data should, therefore, be treated with caution. The descriptors for each band (from 'new to English' to 'fluent') are provided in Appendix K.

[16] For instance, a minimum of 14 eligible Year 10 EAL pupils in science and 14 eligible Year 10 EAL pupils in history. There could also be overlap between the two samples (i.e. a participating Year 10 science student could also be taking GCSE history and be included in both target recruitment numbers).

was recalculated with an estimated MDES of .31[17] at randomisation. This means that the effect of the intervention measured in (adjusted) between-group differences in the outcome variables needs to be larger than originally anticipated (MDES = 0.20, R^2 = 0.010 compared to MDES = 0.31, R^2 = 0.015), which is also discussed in more detail in the 'Limitations' section of the report.

Both the EEF and the evaluation team conducted further analyses taking the variation in school sizes into account and came to an agreement that the associated uncertainty in the MDES was acceptable. All MDES including the MDES at analysis stage are presented in Table 9.

Since the study is focused on a specific sub-population (EAL pupils), performing further subgroup analyses with satisfactory levels of statistical power is difficult. Due to the small numbers in the primary population (EAL pupils) a subgroup analysis for FSM status can only provide broadly indicative results regarding the effect in this specific subgroup. At the time of the trial, to our knowledge no estimate of the share of Year 10 pupils who fulfil both FSM and EAL criteria existed. We therefore, expected conservatively about 20% of the EAL pupils per school to fulfil both criteria. For our 71 recruited schools with three FSM+EAL pupils each, the MDES for this sub-population was estimated at .38 at randomisation (see also Table 9). This means that the intervention would have needed to be more effective in this subgroup than planned (the intervention effect needs to lead to a stronger signal in this smaller sample) to be detectable. A problem that is acknowledged for such subgroup analyses (Inglis *et al.,* 2018; Petticrew *et al.,* 2012).

Randomisation

Schools were only eligible for randomisation after:

- the headteacher signed a Memorandum of Understanding;
- teacher consent was obtained;
- pupil details for the trial teachers (UPNs, pre-specified demographics,[18] 'EAL in the Mainstream Classroom' designated evaluation teacher and subject details) were provided; and
- completion by all participating teachers of a pre-randomisation teacher survey.

Group allocation was conducted at the school level using minimisation. Minimisation uses algorithms to ensure balance at baseline and permits ongoing allocation, so schools know which condition they have been assigned to soon after recruitment. The 12 Delivery Centres (i.e. geographical region) were used to stratify the allocation. This was in order to take into account the capacity of the Delivery Centres to deliver training and to ensure comparability within each Delivery Centre region. Randomisation was conducted by the evaluation team using MinimPy software (Saghaei and Saghaei, 2011) between 29 September 2017 and 18 December 2017. A total of 71 schools were randomised in batches (29/9/17, 38 schools; 6/10/17, 6 schools; 15/11/17, 8 schools; 20/11/17, 6 schools; 29/11/17, 3 schools; 14/1 2/17, 8 schools; and 18/12/17, 2 schools): 33 to the control condition and 38 to the intervention condition, with 2,421 pupils in the control condition and 2,413 in the intervention condition.

Statistical analysis

All analyses have been conducted by Jan R. Boehnke (University of Dundee). All data is presented descriptively with means and standard deviations (SDs) for quantitative outcomes and category frequencies for categorical data. All statistical analyses are reported for complete cases; sensitivity analyses for missing data were planned for the primary outcome, but since missingness and dropout amounted to less than 5% for that set of analyses, none were conducted (see **Statistical Analysis Plan**). Bootstrapped confidence intervals (CIs) were used to judge the statistical significance of the intervention effect. A CI offers information about the uncertainty associated with a point estimate from a given

[17] Since the inception of the project, the EEF guidance on statistical analyses has changed. Originally it was envisaged to control for school-level variation in achievement in the analyses. This would have boosted the trial's statistical power but would have made generalisation of the results more difficult (since they would have been conditional on between-school variation in prior achievement). When the changes for the cohort were agreed, it was also agreed in line with the statistical guidance to no longer control for between-school variation in the analysis. The MDES with the original assumption of a between-school correlation of r = .50 and with the realised sample size in Cohort 1 would have been 0.27.

[18] These were: first name, last name, and date of birth for NPD matching; EAL binary descriptor (Yes/No); and EAL proficiency in English level.

sample (Gardner and Altman, 1988). Briefly, assuming assumptions hold (Greenland *et al.*, 2016), a 95% CI describes the range of values in which 95% of replications would fall, if the study was repeated. For example, out of 100 replications of our trial, 95 of the estimated effect sizes would include the true effect. When the CI includes values that represent no difference between the two trial groups (e.g. an effect size of zero), it means that, given the current sample and analysis, we could not exclude it as a potential result (i.e. insufficient evidence to reject the statistical hypothesis of 'no difference' between the trial groups). As the sample size calculation for the trial was conducted for a pre-defined effect size that we minimally aimed to detect (MDES = 0.20), the CI allows us to evaluate whether the pre-defined effect size is a likely outcome with reference to this range of estimates. Additionally, there is a straightforward connection to statistical significance tests as any value outside that CI would lead to a p-value of $p < 0.05$ (Greenland *et al.*, 2016), which is relied on for the interpretation of the primary and secondary outcome analyses (see criteria below). Bootstrapped CIs take into account violations of normality assumptions in model error distributions (Bland and Altman, 2015) and since they are recommended to evaluate the variation in mediation and cross-level interaction terms in our analyses (e.g. Pituch *et al.,* 2006), we applied them consistently throughout.

Primary analysis

The primary research question of this trial is: How effective is the 'EAL in the Mainstream Classroom' programme in improving subject-specific academic attainment when delivered to Key Stage 4 EAL pupils taking GCSE science? For this analysis only EAL pupils were included and a mixed effects model in which pupils are nested within schools was used. This made it possible to separate within-school variation in the outcome from between-school variation.

The analysis is intention-to-treat (ITT), which means that schools were treated according to the condition they were allocated to (control or intervention), not that which they actually received. The analyses used cluster-bootstrapped CIs (e.g., Huang, 2018) to account for potential violations of distributional assumptions. From each school, a random sample of the same size as its actual sample was drawn (with replacement) and across these school-wise bootstrap samples, the mixed model was then estimated.[19] This process was repeated b = 1,000 times and for a 95% CI, the statistical estimates were saved, and their top and bottom 2.5% quantiles identified. The average of the bootstrapped values was treated as the point estimate and reported in all coefficient tables. No p-values are reported.

This study was planned for a single primary outcome, the Key Stage 4 science GCSE. In accordance with the power analysis, pre-test data from the Key Stage 2 SATs reading raw scores (KS2_READMRK, possible range 0–50) was used as the pupil-level covariate without random variation across schools. An individual student *i*'s KS4SCI (see justification above; see also Table 5[20]) result in a specific school was modelled as depending on school *j*'s random school-level intercept (μ_{0j}) and a pupil-level error term (ε_{ij}). Each school's random intercept (μ_{0j}) depends on an overall intercept (average performance; γ_{00}); each school's level on the stratification variable, which controls for the Delivery Centre (DC); the intervention to which the school was randomised (CP, testing for the intervention effect) and contain a school-level error term (u_{00}):

$$KS4SCI_{ij} = \mu_{0j} + \mu_{1j}KS2_{ij} + \varepsilon_{ij} \quad (1)$$

$$\mu_{0j} = \gamma_{00} + \gamma_{01}DC_{0j} + \gamma_{02}CP_{0j} + u_{00} \quad (2)$$

$$\mu_{1j} = \gamma_{10} \quad (3)$$

$$\mu_{2j} = \gamma_{20} \quad (4)$$

The analysis was performed in the R environment (R Core Team, 2020); specifically, the R-package lme4 (Bates *et al.,* 2015) was used with the corresponding formula expression in the command lmer():

[19] For example, if there were observations 1,2,3,4,5 in a school, one resample could be [1,2,2,5,4] and another [1,5,1,1,3].

[20] This variable is a composite of two calculations. First, the grade achieved in full GCSE combined science (KS4_APCOMBSCI_91) was read out where documented, and split into two variables, each presenting one GCSE score for a science subject. Values 1–9 were treated as acceptable data and others were treated as missing (e.g. missing scores or unmarked results). These two variables were then averaged, resulting in a single average value if both scores were present, or the achieved score in one of the two subjects if only one was documented. Second, an average of the three science subjects (biology, chemistry, and physics; KS4_APBIO_91, KS4_APCHE_91, and KS4_APPHY_91, respectively) was calculated following the same approach (e.g. an average of two scores if only two were documented). The variable KS4SCI then takes on the value of the first variable for those pupils taking the combined science classes; and the value of the second for those pupils taking three sciences.

KS4SCI ~ KS2 + DC + CP + (1 | school)

The intervention would be evaluated as potentially effective in this trial when the average bootstrapped point estimate for the coefficient of the intervention effect (γ_{02}) reflected that intervention schools achieved on average higher scores in GCSE science attainment (KS4SCI) and the 95%-bootstrap CI of this coefficient does not include zero.

Secondary analysis

The secondary outcomes are: How effective is the 'EAL in the Mainstream Classroom' programme in improving subject-specific academic attainment at Key Stage 4 in a second GCSE subject (history) (RQ2); and How effective is the 'EAL in the Mainstream Classroom' programme in improving academic attainment in English (as measured by GCSE English Language) when delivered to Key Stage 4 EAL pupils (RQ3)?[21] The analytic approach used exactly the same procedure and model as for the primary outcome, with one difference: instead of GCSE science attainment (KS4SCI) the secondary outcome variables were used as the dependent variables (GCSE history score; GCSE English language score; see above).

As before, the intervention was evaluated as having shown a potential effect on a secondary outcome when the 95%-bootstrap CI of the coefficient (γ_{02}; see formula 2 above) did not include zero. This result cannot be used to gauge the efficacy of the intervention and is reported purely for exploratory purposes to evaluate whether there are potential positive or negative effects of the intervention on curriculum outcomes, which would need further research.

Analysis in the presence of non-compliance

To assess compliance, attendance at training was used as a proxy measure. Given that there were three training sessions this was on a scale of '0' (no training attended, 0/3 sessions) to '3' (all three sessions attended, 3/3 sessions). To analyse the effect of compliance on the primary outcome of EAL pupils, the approach outlined by Steele *et al.* (2007) was used. In this approach, two multilevel models are estimated; one is the same as the primary outcome analysis (e.g. 1–4), except that the intervention variable is replaced by the compliance measure; in the other, the compliance measure is predicted by the intervention group and additional variables (see below). The random effects across these equations are allowed to correlate in the simultaneous estimation procedure to capture characteristics that influence either dependent variable:

- equation system 1 will be the same as equations 1–4 with the respective compliance variable replacing the intervention variable (CP); and
- equation system 2 will have the respective compliance variable as the dependent variable (treating it as a student-level variable since both may vary within one school), predicted by the intervention (CP) and other variables specified in equation 1–4 above; additionally the schools' averages of pre-treatment Key Stage 2 SATs reading raw scores, FSM, and proficiency in English (taking all pupils into account) will be used as predictors, to proxy potential selection effects due to better performing schools and/or pupils being drawn from higher SES backgrounds.

The intention was to treat the compliance variable as categorical. However, this was not possible with the available software in the Secure Research Service. Therefore, the compliance variable was treated as a continuous indicator. The analysis was conducted using MLwiN (Version 3.04, Centre for Multilevel Modelling, University of Bristol, UK; Charlton *et al.,* 2019). The procedure results in an estimate of the relationship between the compliance variable and the primary outcome (science score, KS4SCI), corrected for potential selection effects as captured by the second equation. The results of both equations and random effects are reported and the results discussed. We also evaluated how plausible it was that one or more variables serve as an 'instrument' (i.e. predicting compliance, but not science outcome; Steele *et al.*, 2007). The intention was again to bootstrap these results, but despite considerable changes to the estimation environment since the inception of the project, this was not possible as the bootstrap method for bivariate clustered outcomes continues to be under development. The interpretation focused on the interpretation of the ratio between coefficient and estimated standard error (SE) (Steele *et al.*, 2007).

Missing data analysis

[21] RQs 4 and 5 are addressed in the section on 'Additional analyses'.

The amount of missing data is documented for each variable individually as well as for the patterns of missing values occurring. Further, the relative frequency of pupils with any missing data was also evaluated by school and by treatment arm and to explore potential patterning of missing data across variables, missingness patterns were explored visually via pattern plots. Deviating from the Statistical Analysis Plan (May 2019), detailed results of these analyses could not be presented, as the low counts within schools are below the Secure Research Service reporting threshold for statistical disclosure control (SDC) and we present a summary instead.

Imbalance at baseline

The minimisation and the Statistical Analysis Plan considered only two variables: the Delivery Centre at school level and the Key Stage 2 reading score. As the former included low numbers per Delivery Centre, we only state here that from each Delivery Centre a one:one split of schools across conditions was attempted and that the algorithm allocated at least one school from each Delivery Centre to each arm of the trial.

To investigate imbalance at the pupil level, Key Stage 2 reading score results were presented with means, SDs, and medians; the control and intervention group distribution was displayed via density plots; imbalance evaluated in standardised mean differences of Key Stage 2 reading scores. Category frequencies for gender, EVERFSM_6_P_SPR19 and proficiency in English bands are presented and imbalance were evaluated via effect sizes (Cramer's V; Faul *et al.,* 2007) for cross-tabulations by intervention group. For any imbalances detected for any of the three categorical variables, a robustness analysis was re-run for primary and secondary outcome analyses, including variables showing imbalance as student-level covariates (Key Stage 2 reading score was included as a student-level covariate in all analyses as per power analysis).

Subgroup analyses

Two subgroup analyses were conducted that were identified by the EEF and project partners: the impact of the programme on EAL pupils eligible for FSM; and the impact of the programme on EAL pupils of different proficiency in English bands at the start of the programme. The trial was not powered for either subgroup analysis, therefore these analyses only provide descriptive insights into potential subgroup effects.

The research assessed the impact of the programme on EAL pupils eligible for FSM (EVERFSM_6_P_SPR19). First, the result for the primary outcome is presented descriptively for the EAL+FSM pupils only. Second, to evaluate whether there is a differential effect for FSM pupils, the mixed effects model described for the primary outcome was extended by adding FSM and an interaction term between FSM and the intervention variable (CP) and run on all EAL pupils. The intervention was evaluated as showing a subgroup effect for FSM when the bootstrapped 95% CI for the coefficient for the interaction term did not include zero. As before, this analysis is purely exploratory and does not estimate the efficacy of the intervention itself.

The previous analytic strategy was extended to include a pupil-level covariate for FSM, which has a random effect across schools and this variation was predicted by the intervention variable:

$$\boldsymbol{KS4SCI_{ij}} = \boldsymbol{\mu_{0j}} + \boldsymbol{\mu_{1j}KS2_{ij}} + \boldsymbol{\mu_{2j}FSM}_{\mathrm{ij}} + \boldsymbol{\varepsilon_{ij}}$$

$$\boldsymbol{\mu_{0j}} = \boldsymbol{\gamma_{00}} + \boldsymbol{\gamma_{01}}\mathrm{DC}_{\boldsymbol{0j}} + \boldsymbol{\gamma_{02}CP_{0j}} + \boldsymbol{u_{00}}$$

$$\boldsymbol{\mu_{1j}} = \boldsymbol{\gamma_{10}}$$

$$\boldsymbol{\mu_{2j}} = \boldsymbol{\gamma_{20}} + \boldsymbol{\gamma_{21}CP_{0j}} + \boldsymbol{u_{20}}$$

The analysis was performed in the R environment with the corresponding formula expression in the command lmer():

KS4SCI ~ KS2 + FSM + DC + CP + CP:FSM + (1 + FSM | school)

The intervention was evaluated as having shown a potential interaction with the specified subgroup variable when the 95%-bootstrap CI of (γ_{21}) did not include zero.

The approach for testing the effect for different baseline fluency levels was similar. Again, only EAL pupil data was analysed. The fluency measure was entered effect-coded into the regression equation. Since the programme is targeted at the mid-fluency levels (i.e. not A's and E's), it was originally planned to use fluency level B as a reference category.

Nevertheless, only very few EAL pupils in the sample were assigned fluency levels of C or lower (with the majority being rated 'E' for fluent). Consequently, only a test of lower (A–C) versus higher (D–E) levels of attainment was possible, and even in this case the results need to be interpreted with additional caution as some schools did not, even with this dichotomisation have any pupils with lower fluency levels. The corresponding formula expression for the random effects model in the command lmer() in R is:

KS4SCI ~ KS2 + fluDu + CP + CP + CP:fluDu + (1 + fluDu | school)

If the CI of a coefficient did not include zero it was evaluated as having a potential (differential) effect. In that case, the interaction is also visually displayed to help interpretation.

Additional analyses and robustness checks

Three additional analyses were performed.

First, the impact of 'EAL in the Mainstream Classroom' when pupils experienced the approach from more than one teacher in more than one subject area (i.e. when pupils were taught by 'EAL in the Mainstream Classroom' intervention teachers in both science and history GCSE subjects) was analysed (RQ4). The analytic approach for the primary and secondary outcomes relating to the impact of the programme on science, history, and English language GCSEs was extended by adding a pupil-level predictor to the equation that captured how much pupils were exposed to the programme. For this we coded, at pupil level, whether pupils were taught by 'EAL in the Mainstream Classroom' intervention teachers in more than one discipline (versus only in one discipline).[22] We re-ran the same model as for the secondary outcome analysis with GCSE Key Stage 4 English language as the dependent variable and added this exposure score as a pupil-level covariate. Adding this coefficient allowed us to test whether being taught by multiple teachers trained in the programme across two disciplines had an additional effect over and above programme provision in itself, for instance, if it impacted on a wider academic attainment outside of science and history. If the bootstrapped CI of the covariate's coefficient did not include zero it was evaluated as having a potential additional effect and the direction of the effect was interpreted.

Second, we tested whether an EAL student's choice regarding combined science/three separate science GCSEs introduced heterogeneity in the programme's effect due to possibly different profiles of pupils taking these options. For this we coded at the pupil level, whether they chose combined science or the three science GCSEs. As it is plausible that there is a difference between the two groups in terms of performance in science GCSEs (and be it only due to the two different exams), we included over and above the pre-specification in the Statistical Analysis Plan also an interaction effect with the treatment to be able to separate a difference in exams from a difference in effectiveness of the intervention. We re-ran the same model as for the primary outcome analysis and the added coefficient for combined science/three separate science GCSEs, allowing us to test whether the choice of combined science classes compared to three separate science GCSEs has an effect on the primary outcome.[23] If the bootstrapped CI of the covariate's coefficient did not include zero it was evaluated as having a potential additional effect and the direction of the effect was interpreted.

Third, we tested the impact of the 'EAL in the Mainstream Classroom' programme on non-EAL pupils within the same classrooms (RQ5). We: (a) used all available student data, i.e. EAL and non-EAL pupils; and (b) added a cross-level interaction effect between EAL/non-EAL and the programme. These coefficients assessed whether there was an effect of the intervention across EAL and non-EAL pupils, and whether this effect was differential between the two pupil groups (e.g. higher averages in the EAL population). If the CI of the coefficient did not include zero it was evaluated as having a potential differential effect. In that case, the interaction was also visually displayed to help interpretation.

Estimation of effect sizes

Effect sizes were calculated based on the total variance in the models. For descriptive differences Hedges g was used as suggested by the EEF guidance. For the results from our mixed-model analyses, the estimated intervention effect as

[22] The original text stated we would code how many teachers of their classes attended the training. The hypothesis to be tested was nevertheless not only about the number of teachers, but also whether they were taught in more than one discipline. This approach reflected this question better, given that schools had also multiple teachers teaching within one discipline.
[23] A small number of control schools within the sample offered only combined science and not the three separate science GCSE option to their students (n not disclosed).

represented by the intervention's coefficient in a model (Effect) was divided by the total variance in the model (see definition of error terms above): $ES = \frac{Effect}{\sqrt[2]{u_{00}+\varepsilon_{ij}}}$

CIs were bootstrapped. Here, *Effect* is the coefficient from the estimated model (e.g., γ_{02} in the analysis of the primary outcome; formula 2).

Estimation of ICC

ICCs were calculated at school level via variance components from a mixed model without any predictors (see section 'Primary outcome analysis'). CIs were calculated via clustered bootstrap.

Implementation and process evaluation (IPE)

Research methods

A light touch IPE was envisaged to minimise the burden on participating schools whilst enabling the evaluation team to inform and explain the impact evaluation outcomes. Given that the primary outcome was GCSE science the IPE solely focuses on science teachers and their Year 10 pupils. The process evaluation was designed to answer the following key research questions:

1) To what extent was the programme implemented with fidelity and what modifications were adopted?

2) What is the impact of 'EAL in the Mainstream Classroom' on the classroom experience?

Given that 'EAL in the Mainstream Classroom' is not a manualised programme, rather a programme of CPD for teachers to adapt flexibly into their classroom practice, the aim was to assess to what extent the programme became embedded into practice and how the strategies taught by the programme were used through triangulation of the survey, interview, and observation data. The small number of observations (detailed above) meant that a case study analysis was no longer deemed appropriate to answer the second research question. Consequently, the revised IPE guidance from the EEF (Humphrey *et al.,* 2019), alongside a more focused analysis based on the logic model was used to structure the analysis of the IPE (which was not pre-determined in the protocol) to answer both research questions. The aspects of the logic model and implementation dimensions from the revised IPE guidance deemed most appropriate as foci for analysis by the evaluation team were as follows:

- The reach of the programme i.e. schools, teachers (including emerging leadership status as specified in the logic model), and pupils (in terms of EAL profile and ability levels) participating in the evaluation. Within this we also included the extent to which pupils were taught by 'EAL in the Mainstream Classroom' intervention teachers (given that the adaptable nature of the programme meant that the amount of exposure to the 'programme strategies' could not be quantified).
- Implementation fidelity, which given the adaptable nature of the programme was related to teachers' decision-making in choosing how to use the programme strategies in their classroom practice and the extent to which they felt the programme influenced their teaching.
- Dosage, which was of interest given that the programme was designed to be continued in the second year of the evaluation (i.e. when the Year 10 pupils were in Year 11), which would give a further period of embedding.

In addition, the IPE explores science teachers' experience of training, and teachers' learning, given their centrality in the logic model. This is done primarily through the teacher survey, with richer detail from the interviews and supplemented by monitoring data from Challenge Partners. The 'business as usual' model in schools was also explored via the IPE by using baseline science teacher survey data from both intervention and control schools and end of the first evaluation year data from control school science teachers.

Further details on the forms of data collection are provided below.

Monitoring information: Given that this was an efficacy trial we anticipated a high level of monitoring and mentoring with schools to be conducted by the developers. The evaluation team worked with Challenge Partners during the additional development phase (i.e. during the pilot year) to maximise the utility of the monitoring data for the needs of the evaluation as well as the developers. The following data was shared with the evaluation team: attendance at training

sessions; post-workshop evaluation forms; and pre-workshop and between workshop (intersessional) teacher task completion records. These were used to provide context for the extent to which teachers engaged with the training (through attendance and completion of tasks).

Teacher survey: An online teacher survey was administered to all intervention and control teachers in the primary subject specialism (i.e. science) using Qualtrics software (Qualtrics, 2015) at two timepoints; baseline (timepoint 1 [T1]; Autumn Term, 2017) and follow up 1 (T2; Summer Term, 2018). The survey was designed to provide the evaluation team with a comprehensive picture of school, teacher, and student context (reach), the teaching of EAL pupils and school and class context in the trial schools (business as usual) alongside understandings, adaptations and experience of the programme and training (intervention-only schools; implementation, adaptations), any other strategies or programmes used by teachers in the trial, and teacher profession knowledge and experience (business as usual). A further follow-up survey was administered to intervention science teachers at the end of the second year of the trial i.e. when the pupils would have been in Year 11 (T3; Summer Term, 2019) to explore possible continued use of the programme and any longer term potential impacts of the CPD. The surveys were developed drawing on existing team knowledge of provision for EAL pupils and of the programme garnered through the pilot year and contained a mixture of closed- and open-ended questions. Copies of the surveys for each timepoint that were administered are provided in Appendix L. Of the 116 (60 intervention, 56 control) science teachers participating in the evaluation, the survey was completed by 106 science teachers at T1 (60 intervention, 46 control; 92% completion rate),[24] 73 science teachers at T2 (42 intervention, 31 control; 63% completion rate), and 38 intervention teachers at T3 (63% completion rate).

School visits: School visits were planned to be carried out at eight intervention schools during the latter part of the Summer Term 2018, at the completion of the intervention. They were designed to encompass an interview with one of the evaluation science teachers, together with a science lesson observation. Schools were randomly chosen from six Delivery Centres to ensure there was a spread of training experiences (given that training was co-delivered by the Delivery Centres). The interviews and observations were designed to inform the evaluation about pupil and teacher engagement with the 'EAL in the Mainstream Classroom' intervention and the translation of the CPD into the classroom context. Although eight school visits were conducted, two of these schools did not sign the GDPR-compliant addendum to the Memorandum of Understanding and their data had to therefore be subsequently withdrawn, meaning that data from only six school visits was subsequently analysed.

Science lesson observations: These were conducted by three researchers, with two visits being conducted by two researchers together to ensure a shared understanding and for quality assurance purposes (i.e. to check for inter-rater reliability). Where there was disagreement, discussion was held between all three researchers until agreement was reached. The observations consisted of two observer checklists designed to capture:

1) General classroom behaviours as the strategies are meant to be embedded into normal classroom practice rather than taught separately. Observers were looking for classroom behaviours, by both teachers and pupils. The items were chosen to assess whether or not this was usual practice for a lesson ('pupils are familiar with tasks') and to assess teacher and pupil engagement (with particular regard to teacher knowledge and working patterns [e.g. differentiation, engagement with whole class or small groups] and student confidence). In addition, there were two extra items: 1) 'tasks assigned to pupils are context embedded', which suggests that the work relating to the programme, for example, the use of verbs in sentence structure was embedded within the context of the science topic being covered; and 2) visual aids used. This latter item is encompassed within the training in order to encourage ordering and organisation of written answers, and is also a common feature within science classes more generally.
2) Identify which areas of learning were focused on within the lessons. These activities were all associated with strategies and techniques emphasised within the 'EAL in the Mainstream Classroom' CPD.

Finally, the researchers made detailed notes on the lessons and the exact strategies used in the lessons. The lessons were then graded as either at 'emerging', 'establishing', or 'embedded' levels of implementation of the programme in the classroom. The observation schedule can be found in Appendix M.

[24] Whilst 60 teachers later allocated to the intervention group were participating in the evaluation at baseline, four subsequently withdrew from the evaluation after randomisation, only one of whom was replaced by another science teacher within the school. Their data was, however, retained for analysis purposes.

Science teacher interviews: These asked questions relating to school, teacher, and student context, reasons for taking the 'EAL in the Mainstream Classroom' CPD, experiences of the CPD, and adaptation and implementation in the classroom. Where possible, the observations (which preceded the interview) were also discussed. One school was unable to facilitate the observation on the day of the researcher visit, although an interview was possible. A further school visited for an observation subsequently did not complete the GDPR-compliant data sharing agreement and therefore we are unable to report on the observation and interview undertaken. Consequently, six science lesson observations took place and eight interviews, with two science teachers being interviewed in one school. Three researchers conducted the visits. Two of the school visits involved two researchers (organised at the start of the school visits) to ensure a shared understanding of implementation of the programme and for quality assurance purposes.

Delivery Centre interviews: Eight telephone interviews were conducted with operational staff in six of the Delivery Centres to understand the training model in practice as well as their views on how effectively the training was delivered. This aspect of the IPE was not included in the protocol but added as a result of teacher interviews relating to school recruitment and attendance at training.

In addition, two researchers attended two training days each in order to better understand the programme, the CPD, and assist in the design of the IPE instruments, although this was not included in the protocol.

Analysis

Both quantitative and qualitative analysis of the process evaluation data has been conducted. The surveys were downloaded into SPSS (SPSS Inc., Chicago, IL, USA) and used to provide descriptive statistics, as well as enabling comparisons to be made both within surveys (i.e. cross-sectional, comparing control and intervention using crosstabulation, and across surveys [again, using crosstabulation] comparing responses across timepoints [T1 and T2 across intervention and control and T3—intervention only—to gather information on continued implementation in the second year of the evaluation]) to assess balance at baseline in existing practice and context and changes over time, including perceived changes in student understanding, confidence, and engagement. Responses are usually presented at the individual level to assess teacher change although in some cases (e.g. when assessing changes in school-level support or to describe participating classes) responses have been collated and are reported at the class level or school level. Where this has been done it is clearly stated in the findings. Proportions in the descriptive data only include those who responded to the item and the sample size (n) is presented alongside.

The interviews were transcribed professionally and downloaded into NVivo software (Lumivero, 2010). Given that the IPE was designed to answer specific questions and assess the logic model a deductive approach was taken (Bingham and Witowsky, 2022). This used the two principal research questions relating to the IPE and the logic model as a framework within which to assign codes, namely; dosage, compliance, experience of training, teacher learning, and implementation fidelity (including adaptations). Within these themes and questions additional codes were added inductively to allow the data to add additional facets to the researcher's understandings of the data. These codes were applied to the interview data and, subsequently, to the free-text responses in the survey questions. A shorter version of this coding frame was applied to the Delivery Centre interviews covering reach, compliance, and experience of conducting the training (where applicable). All coding of data at the individual level was undertaken by one researcher to ensure reliability and consistency (i.e. the Delivery Centre interviews were coded by one researcher and the teacher interviews by a second researcher in consultation with the project lead).

Data from the observation schedules was input into Excel and analysed descriptively. Alongside this, the researcher notes were typed up and analysed by hand to provide a narrative analysis of the main points relating to the lesson in relation to the 'EAL in the Mainstream Classroom' and the extent to which they were embedded within the lesson. A holistic assessment was then made by the researcher on the extent to which the programme was embedded in that particular lesson (i.e. this snapshot of classroom practices), which was discussed with other researchers on the evaluation team to ensure that a shared understanding was reached. These are presented as vignettes in the findings section below in order to answer: What does 'EAL in the Mainstream Classroom' 'look like' in the classroom? The use of a case study approach was posited in the protocol but given the small number of schools visited and interviews conducted, it was decided that to do so would not be sufficiently pseudonymised. This means that some of the nuances connecting training into practice may have been lost from the evaluation. Detailed descriptions of the observations from the researcher's notes are presented in the IPE findings, supplemented by relevant data from the teacher interviews. These should not, however, be subject to wider generalisation or interpretation.

In addition to analysing the data separately, we have triangulated the findings to inform the impact analysis and to understand the outcomes of the evaluation more fully.

Table 6 provides a summary of the IPE data collected, the form of analysis that was conducted, and their relationship to the research questions and logic model.

Table 6: IPE methods overview

Research methods	Data collection methods	Participants/ data sources	Data analysis methods	IPE research questions addressed	Implementation dimensions
Monitoring data	Training monitoring data	Collected by Challenge Partners and Delivery Centres	Descriptive analysis in SPSS	1	
Survey	Science teacher survey	Baseline (T1), post-test (T2) and follow-up (T3, intervention only, science teachers in participating schools	Descriptive and crosstabs analysis in SPSS	1, 2	Reach, dosage, experience of training, teacher learning, implementation fidelity, adaptations
Observation	Observations	Classroom observation of Year 10 science lesson in intervention schools	Descriptive analysis in Excel. Descriptive, narrative analysis by hand	1, 2	Implementation fidelity, adaptations
Interview	Teacher interviews	Year 10 intervention science teachers	Deductive, thematic analysis in NVivo	1, 2	Dosage, experience of training, teacher learning, implementation fidelity, adaptations
Interview	Delivery Centre interviews	Delivery Centre operational staff (i.e. facilitators, administration)	Deductive, thematic analysis in NVivo	1	Reach

Costs

The researchers sought to assess the initial cost of programme delivery (i.e. training) and implementation within schools. Following the EEF guidance (EEF, 2019), the cost is presented as the cost of the 'EAL in the Mainstream Classroom' programme per pupil per year over a three-year period.

The cost of the intervention, including training and support, distribution of resources and materials during training, provision for workshops, and time required for staff training was collected from the developers. Costs relating to any additional materials or resources needed within schools for the intervention was collected through teacher interviews as part of the process evaluation. This enabled identification of any issues around provision/cost of resources and implementation that may not have been previously identified. Costs were calculated based on number of Delivery Centres (initially 12, reducing to 11 after the first workshop), number of workshops held (three per Delivery Centre), and the average number of pupils in intervention schools taught by participating teachers.

Timeline

Table 7: Timeline

Dates	Activity	Staff responsible / leading
Summer–Autumn Terms 2017	Recruit schools	Delivery Centres, Challenge Partners, University of York
October–December 2017	Randomise	University of York
Summer–Autumn Terms 2018	Collect school data (including unique pupil numbers and fluency indicators) and science teacher surveys (T1)	University of York
Spring–Summer Terms 2018	Training	Challenge Partners with Delivery Centres
Spring–Summer Terms 2018	Researchers attend training	University of York
Spring–Summer 2018 (and academic year 2019/2020)*	Teachers deliver intervention in Year 10 (and potential subsequent implementation in Year 11 classrooms)	Schools
Summer Term 2018	Conduct teacher interviews and classroom observations	University of York
End of Summer Term 2018	Collect end of intervention teacher surveys (T2)	University of York
End of Summer Term 2019	Collect follow-up teacher surveys (T3)	University of York
Autumn 2019	Interview Delivery Centres	University of York
Autumn 2020**	Apply for NPD data**	University of York
Spring 2022	Analysis	University of York/University of Dundee
May 2022	Draft report	University of York/University of Dundee
February 2023	Publication of main report	Education Endowment Foundation

* *It was expected that teachers would continue teaching the same classes and using the intervention strategies in Year 11 although this was not a mandatory condition of inclusion within the trial.*

** *Originally it was intended that NPD data would be available from autumn 2019. However, changes in access to NPD data and the following pandemic from March 2020 meant that all activities designed to take place from autumn 2019 were substantially delayed. This timeline details the actual timing of events as opposed to those presented in the study protocol.*

Impact evaluation results

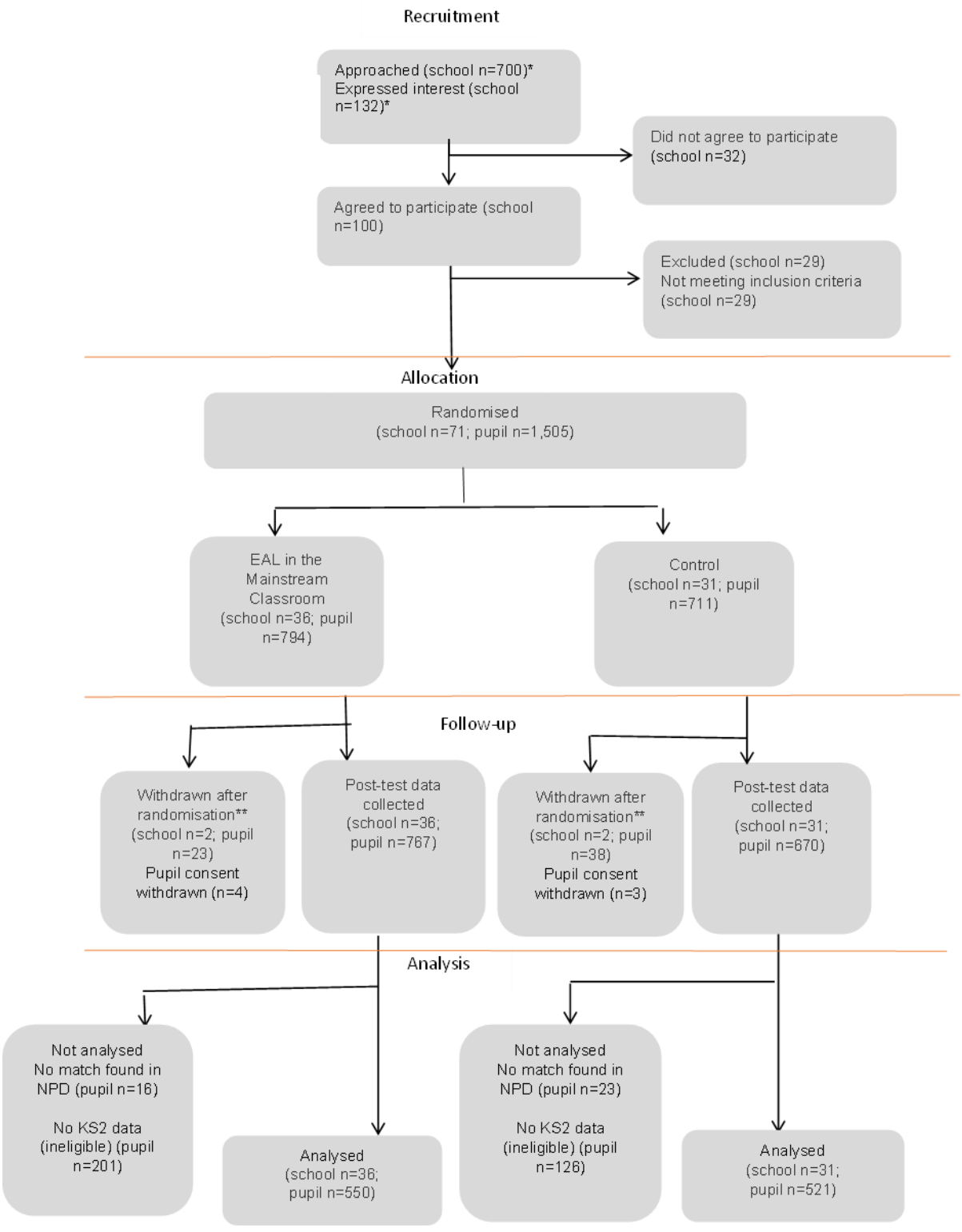
Participant flow including losses and exclusions

Given that Delivery Centres recruited schools from the lists provided by Challenge Partners and from their own personal contacts and networks we are unsure of the total number of schools approached to take part in this trial. However, the evaluation team received 132 expressions of interest. Of these, 32 did not progress to signing a Memorandum of Understanding and 29 were excluded as ineligible due to having insufficient EAL pupils in their science or history classes (i.e. less than 12 Year 10 EAL science pupils or less than 12 EAL Year 10 history pupils). Consequently, 5,340 pupils were randomised of whom 1,505 were EAL pupils taking GCSE science classes with a science teacher participating in the evaluation (i.e. pupils with GCSE science results were eligible for inclusion in the primary outcome analysis). Of these 1,505 pupils, 794 pupils were nested within the 36 schools allocated to the intervention condition and 711 pupils were within the 31 schools allocated to the control condition.

Immediately after randomisation (autumn 2017) one intervention school withdrew from the trial. Subsequently a further three schools implicitly withdrew when they did not sign or return the GDPR-compliant addendum to the original Memorandum of Understanding (dated September 2018). Consequently, in the absence of any such agreement, pupil data for all four schools were deleted. This meant that we had data for 1,437 pupils taking GCSE science within 67 schools (767 intervention; 670 control).

On submission of the database to the NPD, it was found that 39 pupils were unable to be matched despite having provided pupils name, date of birth, and UPN (16 intervention; 23 control). A further 327 pupils were found to be ineligible for inclusion in the trial due to not having a recorded Key Stage 2 SATs result (despite initial checking by the research team with individual schools at the point of recruitment and at the point of collecting individual pupil details). Consequently, data for 1,071 pupils (550 intervention, 521 control) was analysed for the primary outcome (GCSE science). The participant flow for the primary outcome is provided in Figure 3. Figure 4 provides the participant flow for the secondary outcomes (GCSE history and GCSE English language), given that GCSE English language was analysed separately for EAL pupils only and for EAL and non-EAL pupils together, both are included in the participant flow diagram.

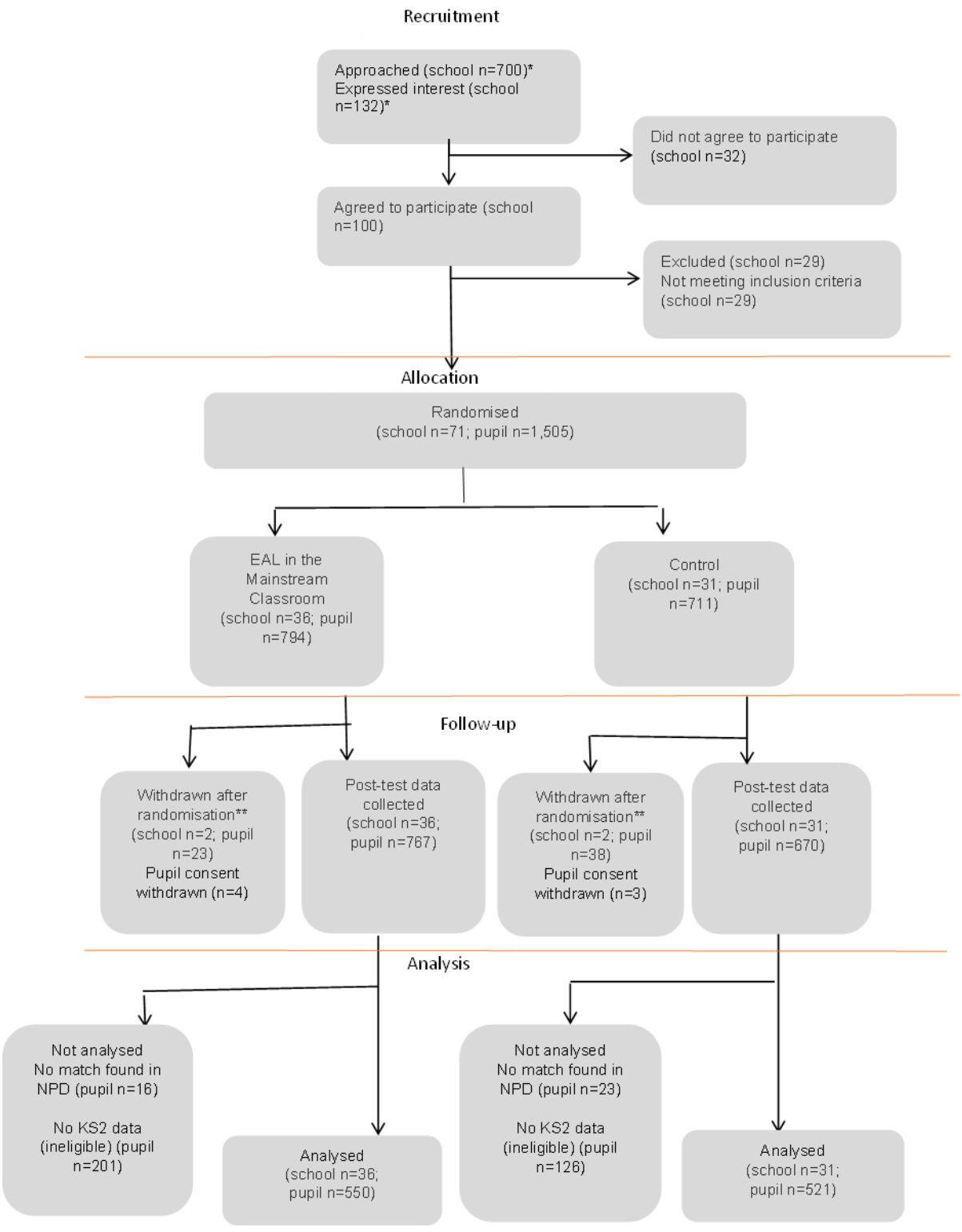
Figure 3: Participant flow diagram – primary outcome (GCSE science)



** Number of schools on the lists provided by Challenge Partners to the Delivery Centres. Given the nature of recruitment (through Delivery Centres) the exact number of schools approached is uncertain.*

*** One intervention school withdrew immediately following randomisation. This school did not complete the GDPR-compliant addendum to the Memorandum of Understanding and neither did a further intervention school and two control schools, which meant that their data could not be used for the purposes of the evaluation.*

Figure 4: Participant flow diagram – secondary outcomes (GCSE history, GCSE English language)



** Number of schools on the lists provided by Challenge Partners to the Delivery Centres. Given the nature of recruitment (through Delivery Centres) the exact number of schools approached is uncertain.*

*** One intervention school withdrew immediately following randomisation. This school did not complete the GDPR-compliant addendum to the Memorandum of Understanding and neither did a further intervention school and two control schools, which meant that their data could not be used for the purposes of the evaluation.*

Attrition

Attrition to the trial for the primary outcome (EAL pupils' GSCE science attainment) is described in Table 8. Overall, 1,505 pupils taking GCSE science within 67 schools (794 intervention; 711 control) were randomised. School withdrawal/failure to complete and return the revised GDPR-compliant addendum mean that four schools withdrew from the study, and an associated 61 EAL Year 10 science pupils (23 intervention, 38 control) were withdrawn from the trial. A further seven pupils (four intervention; three control) were withdrawn individually by their parents within participating schools. All remaining pupil data (relating to 1,437 pupils: 767 intervention; 670 control) was submitted to the NPD for matching. Of these, 39 pupils were unable to be matched (16 intervention; 23 control) and 327 pupils were found to be ineligible for inclusion in the trial due to not having a recorded Key Stage 2 SATs result. Consequently, data from 1,071 pupils were analysed (550 intervention pupils, 521 control). This meant that overall attrition for the trial was 434:1,505 or 28.8%.

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

		Intervention	Control	Total
Number of pupils	Randomised	794	711	1,505
	Analysed	550	521	1,071
Pupil attrition (from randomisation to analysis)	Number	244	190	434
	Percentage	31	27	29

Table 9 presents the MDES at different stages of the project (see explanations in the section on 'Sample Size). Overall, the sample analysed was very similar in size and distribution to what was expected at randomisation without being able to interrogate the NPD data at that point. There are three differences that need to be noted:

- the ICC for the science outcome was markedly higher than originally expected for the primary outcome analyses (and especially in the subgroup analysis for FSM pupils only);
- the number of pupils with EAL status who were taught by a science teacher per school was lower than originally expected (some schools contributed with only one student); this could have contributed to the increased ICC; and
- a higher than originally expected share of pupils eligible for FSM was observed per school, but in two schools no EAL+FSM pupils took part in the project (which does not mean they had none).

These, and the reduced number of schools compared to the randomisation stage, are all factors that contribute to the increased MDES that could have been detected (see discussion of this in the 'Methods' and 'Limitations' sections).

Table 9: MDES at the point of protocol (version 2), at the point of randomisation, and at point of analysis*

		Protocol		Randomisation		Analysis	
		Overall EAL pupils	FSM[1] EAL pupils	Overall EAL pupils	FSM[1] EAL pupils	Overall EAL pupils	FSM EAL pupils
MDES		.22	.31	.31	.38	.44	.46
Pre-test/post-test correlations	Level 1 (pupil)	.50	.50	.50	.50	.58	.56
	Level 2 (class)	–	–	–	–	–	–
	Level 3 (school)	.50	.50	–	–	–	–
ICCs	Level 2 (class)	–	–	–	–	–	–
	Level 3 (school)	.19	.19	.19	.19	.38	.38
Alpha		0.05	0.05	0.05	0.05	0.05	0.05
Power		0.80	0.80	0.80	0.80	0.80	0.80
One-sided or two-sided?		two-sided	two-sided	two-sided	two-sided	two-sided	two-sided
Average cluster size		14	2.8	22.27	4.45	15.99	6.37
Number of schools	Intervention	60	60	38**	38	36	35
	Control	60	60	33**	33	31	30
	Total:	120	120	71	71	67	65
Number of pupils	Intervention	840	168	851	169***	550	234
	Control	840	168	730	147***	521	201
	Total:	1,680	336	1,581	316***	1,071	435

Notes:
See comments on the development of the sample size calculation in the respective chapter of the 'Methods' section.
[1] *Extrapolated based on the assumption of 20% FSM in sample.*
* *Focal numbers are for the primary outcome analysis.*
** *As the randomisation was performed in batches, this resulted in an allocation that slightly favoured the intervention group. It was originally planned to recruit more schools, which would have corrected this imbalance through minimisation.*
*** *Numbers rounded since based on averages.*

Pupil and school characteristics

Characteristics of the 71 schools are presented in Table 10. As randomisation took place by school, school characteristics are presented by allocation alongside the national data. Both sets of data (national data and that relating to participating schools) are taken from 2017 (the year of recruitment to the evaluation). Schools were predominantly from urban areas and with higher-than-average proportions of pupils for whom are EAL, although this was to be expected given the design and recruitment to the trial. They also had higher proportions of pupils in receipt of FSM. In terms of differences between conditions, these trends were more noticeable for intervention schools than control schools. For example, 24% of pupils in intervention schools were noted as being eligible for FSM compared to 22% in control schools, both of which were higher than the 13% national average. Likewise, 49% of intervention school pupils were classed as EAL learners compared to 43% in control schools, both considerably higher than the 16% reported nationally. In terms of attainment, evaluation schools both had lower levels of pupils achieving strong passes in both English and mathematics GCSEs, as a whole and for EAL pupils separately; 43% and 46% nationally for pupils as a whole and EAL pupils, respectively, compared to 35% and 38% for intervention schools, respectively and 37% for both pupils overall and EAL pupils as a group in control schools.

Table 10: Baseline characteristics of groups as randomised

School level (categorical)	National level		Intervention		Control	
	n/N (missing)	Count (%)	n/N (missing)	Count (%)	n/N (missing)	Count (%)
Secondary school type*	3,408/3,408		38/38 (0)		33/33 (0)	
Academies and free schools:						
Converter academy	-	1,471 (43)	-	12 (32)	-	15 (25)
Sponsored academy	-	619 (18)	-	9 (24)	-	12 (36)
Free schools	-	151 (4)	-	0 (0)	-	1 (3)
Other**	-	84 (3)	-	0 (0)	-	
Local Authority Maintained Schools:						
Community School	-	523 (15)	-	9 (24)	-	3 (9)
Foundation school	-	248 (7)	-	6 (16)	-	2 (6)
Voluntary Aided School	-	267 (8)	-	2 (5)	-	0 (0)
Voluntary Controlled School	-	42 (1)	-	0 (0)	-	0 (0)
City Technology Colleges	-	3 (0)	-	0 (0)	-	0 (0)
Urban/rural	3,233/3,408 (175)		38/38 (0)		33/33 (0)	
Rural hamlet/village	-	109 (3)	-	0 (0)	-	0 (0)
Rural town and fringe	-	335 (10)	-	0 (0)	-	1 (3)
Urban city and town	-	1,516 (47)	-	12 (32)	-	12 (36)
Urban major conurbation	-	1,166 (36)	-	23 (61)	-	16 (49)
Urban minor conurbation	-	107 (3)	-	3 (8)	-	4 (12)
Ofsted rating	3,408/3,408 (0)		38/38 (0)		33/33 (0)	
Outstanding		740 (22)	-	5 (13)	-	6 (18)
Good		1,758 (52)	-	25 (66)	-	18 (55)
Requires improvement		558 (16)	-	5 (13)	-	7 (21)
Inadequate		256 (8)	-	1 (3)	-	0 (0)
No Ofsted assessment		96 (3)	-	2 (5)	-	2 (6)
School level (continuous)	**n/N (missing)**	**Mean (SD)**	**n/N (missing)**	**Mean (SD)**	**n/N (missing)**	**Mean (SD)**
FSM eligibility %						
Number of schools	3,408/3,408 (0)	-	38/38 (0)	-	33/33 (0)	-
Percentage of pupils eligible for FSM	-	13.1	-	23.6 (11.2)	-	21.7 (10.6)
EAL						
Number of schools	3,408/3,408 (0)	-	38/38 (0)	-	33/33 (0)	-
Percentage of pupils for whom are EAL		15.8		48.5 (0.2)	-	42.5 (0.2)
School attainment						

Number of schools	3,128/3,408 (250)	-	38/38 (0)	-	33/33 (0)	-
Percentage of pupils achieving strong passes (Grades 9–5) in both English and mathematics GCSEs	-	42.6 (0.2)	-	35.3 (0.12)	-	36.5 (0.1)
Number of schools	2053/3,408 (1325)	-	37/38 (1)	-	33/33 (0)	-
Percentage of pupils for whom are EAL achieving strong passes (Grades 9–5) in both English and mathematics GCSEs	-	45.9 (0.2)	-	38.1 (0.2)	-	37.3 (0.2)
Pupil level* (categorical)**	**n/N (missing)**	**Count (%)**	**n/N (missing)**	**Count (%)**	**n/N (missing)**	**Count (%)**
Gender						
Female			271/1071 (0)	271 (49%)	247/1071 (0)	247 (47%)
Male			279/1071 (0)	279 (51%)	274/1071 (0)	274 (53%)
Eligible for FSM (EverFSM6)			234/1071 (SDC)	234 (43%)	201/1071 (SDC)	201 (39%)
Key Stage 2 reading attainment (KS2_READMRK)			550/1071 (0)	Mean = 28.19 (SD = 8.63)	521/1071 (0)	Mean = 28.58 (SD = 8.87)

** Includes Middle Schools as deemed.*
*** Academy 'other' includes Studio Schools and University Technical Colleges.*
**** Primary outcome pupils only (i.e. Year 10 EAL science pupils).*
Percentages may not add up to 100 due to rounding.
*Sources: **https://www.gov.uk/government/statistics/schools-pupils-and-their-characteristics-january-2017** and Key Stage 4 attainment and school characteristics data extracts, 2017 (see gov.org.uk)*
SDC not disclosed due to statistical disclosure risk; see details in Appendix N, Table 1.

Pupil data were collected at n_S = 67 schools nested in 11 Delivery Centres, with on average 6.09 schools per Delivery Centre (SD = 2.70). Of these schools, n_I = 36 (53.7%) were allocated to the intervention group and n_C = 31 (46.3%) were allocated to the control group. On average, the schools had 64.43 (SD = 24.43) pupils per school and 26.66 (SD = 13.79) EAL pupils.

The descriptive statistics for the student characteristics are displayed in the following table (Table 11), split by research question. For the **primary outcome** analyses only EAL pupils who were taught science by at least one of the project's science teachers were eligible (Table 11, column 3; N = 1,071). For the primary outcome analysis n = 67 schools contributed on average 15.99 pupils (SD = 8.87; min = SDC[25]; max = 46), which is slightly above the expected n = 14 that was specified in the **Statistical Analysis Plan**.

For the secondary analysis **'History'** only EAL pupils who were taught history by at least one of the project's history teachers were eligible (Table 11, column 4; N = 952). To this secondary analysis, n = 65 schools (a reduction by a further two schools in the intervention group) contributed data and had on average 14.65 pupils (SD = 8.73; min = SDC; max = 48).

For the secondary analysis **'English'** only EAL pupils who were taught in any of the two subjects by at least one of the project's teachers were eligible (Table 11, column 5; N = 1,786). To this secondary analysis, n = 67 schools contributed data and had on average 26.66 pupils (SD = 13.79; min = SDC; max = 74).

In addition to the data displayed in Table 11, a small number of pupils were in Year 12,[26] the exact number not being disclosed (SDC); all other pupils were in Year 11.

[25] SDC ('statistical disclosure control') indicates that a frequency was below the minimal frequency defined by the ONS to reduce the risk of identification and was therefore suppressed.

[26] The data are reported as documented in the NPD. The information could have been erroneously reported by the schools or erroneously documented in the NPD. Additionally, pupils could have returned to an earlier year group for a range of reasons. Since pupils were participants as identified by the schools, we prioritised this information and retained them in the analyses.

Imbalance at baseline

Table 11 presents the baseline characteristics of the pupils as randomised. Descriptive details of the distribution of pre-specified key variables across the two-arms of the trial are presented in Appendix N (Table 1). Key Stage 2 SATs reading attainment (KS2_READMRK) scores are presented as means and SDs; and the control and intervention group distribution are displayed via density plots (these plots as well as density plots for the outcomes are also presented in Appendix N).

The investigation for imbalance at baseline showed for the primary analysis set differences in effect sizes that were very small (Appendix N, Table 1). For the secondary analyses, imbalance evaluated in standardised mean differences shows that the samples for the history and English GCSEs may be based on samples not fully balanced across groups, with pupils in the control group showing on average higher baseline scores (small to medium effect size). This baseline variable was included as a pupil-level control variable in all analyses as planned.

Gender, and FSM at baseline are presented in Table 11. In addition, category frequencies were calculated for gender, FSM, and Verbal Fluency. Imbalance was evaluated via Cramer's V (Faul *et al.,* 2007) and the findings presented in Appendix N (Table 1). Based on the effect size evaluation, in all samples apart from the sample recruited for the primary analysis the EAL proficiency bands may have been imbalanced, with pupils in the control group being assessed as in higher proficiency in English bands (all small effect sizes); and the sample for the secondary analysis of the GCSE history showed a lower percentage of pupils receiving FSM in the control group.

Missing data analysis

The amount of missing data is documented for each variable individually as well as for the patterns of missing values occurring (see Table 11). Evaluating levels of missingness in terms of information that is available per pupil, shows high completeness of data for all analyses. For the primary outcome analysis, considering EAL proficiency in English band, FSM status, gender, Key Stage 2 SATs reading attainment (KS2_READMRK) at baseline and the composite GCSE science score as relevant variables for the analyses of all EAL pupils who attended science classes, N = 1,034 (96.5%) had complete data (n = 19, 3.5% pupils with any missing data in the intervention group; n = 18, 3.5% with any missing data in the control group).

For the secondary outcome analysis regarding history, considering EAL proficiency in English band, FSM status, gender, Key Stage 2 SATs reading attainment (KS2_READMRK) at baseline and the Key Stage 4 GCSE history result as relevant variables for the analyses of all EAL pupils who attended history classes, N = 880 (92.4%) had complete data (n = 33, 7.1% pupils with any missing data in the intervention group; n = 39, 8.0% with any missing data in the control group).

For the secondary outcome analysis regarding GCSE English language outcomes, considering EAL proficiency in English band, FSM status, gender, Key Stage 2 SATs reading attainment (KS2_READMRK) at baseline and the Key Stage 4 GCSE English language result as relevant variables for the analyses of all EAL pupils who attended classes with any of the project teachers, N = 1,750 (98.0%) had complete data (n = 18, 2.0% pupils with any missing data in the intervention; n = 18, 2.1% with any missing data in the control group).

Considering EAL status, FSM status, gender, Key Stage 2 SATs reading attainment (KS2_READMRK) at baseline and Key Stage 4 English language attainment as relevant variables for the analyses of all pupils, N = 4,203 (97.4%) had complete data (n = 59, 2.8% pupils with any missing data in the intervention group; n = 55, 2.5% with any missing data in the control group).

To evaluate the impact of missing data on the robustness of findings from the ITT analyses of the primary outcome, sensitivity analyses were planned to evaluate the robustness of the results if > 5% missing data for the primary outcome analysis were encountered (i.e. 5% of cases were deleted listwise for that analysis); and a fully conditional specification (FCS) approach (Enders *et al.,* 2018) would have been employed to impute missing values in the sensitivity analysis. This threshold was not met (only the analysis of the secondary outcome, history for pupils attending history classes taught by project teachers, showed a slightly elevated level of missingness with 7.6%), no further analyses were conducted.

Table 11: Descriptive data for the four different analysis sets used in the analyses and the collected data identified by source (data collected by research team in University of York vs drawn from NPD)

	All pupils (N = 4,317)	Primary outcome (only EAL and taught by any of the science teachers) N = 1,071	Secondary outcome (only EAL and taught by any of the history teachers) N = 952	Secondary outcome English (only EAL and taught by any of the project teachers) N = 1,786
CATEGORICAL DATA				
Gender (NPD: *KS4_GENDER*)[1]				
Female	2,225 (52%)	518 (48%)	530 (56%)	923 (52%)
Male	2,092 (49%)	553 (52%)	422 (44%)	863 (48%)
Missing	0 (0%)	0 (0%)	0 (0%)	0 (0%)
EAL status (York: *EAL*)				
Yes	1,786 (41%)	1,071 (100%)	952 (100%)	1,786 (100%)
No	2,531 (58%)	–	–	–
Missing	(0) (0%)	(0) (0%)	(0) (0%)	(0) (0%)
EAL Proficiency in English band (York: *EAL_Code*)				
New to English (A) and Early Acquisition (B)	57 (1%)	-SDC-	-SDC-	57 (3%)
Developing Competence (C)	283 (7%)	214 (20%)[4]	169 (18%)[4]	283 (16%)
Competent (D)	632 (15%)	363 (34%)	358 (38%)	632 (35%)
Fluent (E)	780 (18%)	473 (44%)	396 (42%)	780 (44%)
non-EAL	2,531 (59%)	–	–	–
Not yet assessed (N) and Missing	34 (1%)	21 (2%)	29 (3%)	34 (2%)
Subjects (York: *Subjects*)				
History	1,714 (40%)	–	715 (75%)	715 (40%)
Science	1,872 (43%)	800 (75%)	–	800 (45%)
History and one or two sciences[2]	682 (16%)	237 (22%)	237 (25%)	237 (13%)
Two sciences	49 (1%)	34 (3%)	–	34 (2%)
Missing	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Condition (York: *Treatment*)				
Intervention	2,124 (49%)	550 (51%)	462 (49%)	909 (51%)
Control	2,193 (51%)	521 (49%)	490 (52%)	877 (49%)
Missing	0 (0%)	0 (0%)	0 (0%)	0 (0%)
FSM (NPD: *EVERFSM_6_P_SPR19*)				
Yes	1,665 (39%)	435 (41%)	407 (43%)	741 (42%)
No	2,615 (61%)	636 (59%)	545 (57%)	1045 (59%)
Missing	37 (1%)	-SDC-[5]	-SDC-[5]	-SDC-[5]
CONTINUOUS DATA				
Reading Score KS2 SATs (NPD: *KS2_READMRK*)				
Mean (SD)	30.07 (8.61)	28.38 (8.74)	28.83 (8.41)	28.50 (8.58)
Missing	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Combined Science (NPD, averaged GSCE results from: *KS4_APCOMBSCI_91*)				
Mean (SD)	4.03 (1.63)	4.18 (1.73)	4.22 (1.63)	4.16 (1.67)
Missing	1,507 (35%)	437 (41%)	295 (31%)	630 (35%)
Triple Science (NPD, averaged results from GCSE Physics, GCSE Chemistry and GCSE Biology): *KS4_APPHY_91 KS4_APCHE_91 KS4_APBIO_91*)[3]				

	All pupils (N = 4,317)	Primary outcome (only EAL and taught by any of the science teachers) N = 1,071	Secondary outcome (only EAL and taught by any of the history teachers) N = 952	Secondary outcome English (only EAL and taught by any of the project teachers) N = 1,786
Mean (SD)	5.93 (1.68)	5.85 (1.59)	5.96 (1.68)	5.92 (1.60)
Missing	2969 (69%)	669 (63%)	689 (72%)	1209 (68%)
Sciences score, primary outcome (either the aggregated Combined Science or Triple Science score)				
Mean (SD)	4.65 (1.87)	4.83 (1.86)	4.72 (1.83)	4.74 (1.84)
Missing	159 (4%)	35 (3%)	32 (3%)	53 (3%)
History GCSE (NPD: *KS4HIS*)				
Mean (SD)	4.47 (2.23)	4.73 (2.19)	4.51 (2.11)	4.52 (2.13)
Missing	1490 (35%)	557 (52%)	72 (8%)	605 (35%)
English Language GCSE (NPD: *KS4ENG*)				
Mean (SD)	4.78 (1.78)	4.80 (1.69)	4.87 (1.68)	4.80 (1.70)
Missing	95 (2%)	19 (2%)	18 (2%)	30 (2%)

Percentages may not add up to 100 due to rounding.
Note. '–' No data for this category in this sub-sample due to selection criteria. SDC This field was edited for SDC to compensate for low frequency of responses; whilst in several places in this table the cell frequencies were high enough to comply with standard approaches, due to the combination with the report of baseline imbalance analyses (see Appendix N, Table 1), this would have made smaller numbers identifiable due to using both tables together.
[1]The University of York team also collected data on gender, but the NPD data were used; among the N = 4,317 presented in column 2, n = 27 (0.6%) disagreements were detected between the two data sources. [2]Collated categories due to rarity of one of the subject combinations. [3]Out of those who had at least one score on these three variables (N = 1,348), 20 had one of the three scores missing (average of remaining two used for analysis); and 15 had two scores missing (remaining score used for analysis). [4]Combined frequency for New to English (A), Early Acquisition (B), and Developing Competence (C). [5]Not disclosed due to low numbers; count of missing observations added 50/50 split to the observed categories; if this resulted in an uneven number, the larger number of observations was added to the category with the larger count.

Outcomes and analysis

This section reports the results of the primary and secondary outcome analyses, summarised in Table 12. The columns present first the number of missing observations by group for the respective analysis, as well as the means and their CIs for the target outcome (neither corrected for clustering nor for the specified covariates). The effect size estimates presented have been derived from the statistical analysis, which was described in the section on 'Estimation of effect sizes' above (EEF, 2018, p. 4; see also: Hedges, 2007; Xiao *et al.,* 2016).

The table reports the observed missing responses and totals. Since the analyses were conducted via bootstrap, missing data/the analysed N varies slightly within each run. The tables with the individual model results in Appendix O provide further details (e.g. SDs of the number of selected pupils to gauge the variability). Whilst this procedure could in principle lead to wider CIs for the effect size estimates since sometimes fewer pupils than the available complete responses are selected. Nevertheless, the SDs for the number of selected pupils are small compared to the overall sample size (see details in Appendix O), which indicates that very similar numbers of pupils were analysed in each run and the risk of widening CIs was minimal.

Primary outcome analysis

The primary outcome analysis was conducted as described above and was planned to answer the question, how effective the 'EAL in the Mainstream Classroom' programme is in improving subject-specific academic attainment when delivered to Key Stage 4 EAL pupils taking GCSE science. As presented in Table 12, the adjusted effect size based on the analytic model is small (effect size: 0.06; 95% CI: -0.06 – 0.18; N = 1,034 analysed, see section on 'Missing data' and Table 12 for more detail). Whilst the direction indicates that the intervention group showed on average higher scores (which equates to one month's additional progress), the CI includes '0'. According to the pre-set criteria, we therefore cannot be confident that there was a difference between the groups in this trial. As the imbalance analysis indicated no potential imbalances, no additional analyses were conducted to evaluate potential impact of these.

Due to the substantial changes to the planned sample size and the fairly large MDES presented in Table 9, it is worth noting that the CI for this effect size estimate is fairly narrow (-0.06, 0.18) and it does not include effect sizes in either direction that were originally seen as potentially relevant (which would have been MDES ≥ |0.20|).

Table 12: Primary and secondary analysis results

Outcome	Unadjusted means				Effect size	
	Intervention group		Control group			
	N (missing)	Mean (95% CI)	N (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges g (95% CI)
Primary outcome Science, EAL pupils only	19	4.83 (4.71, 4.95)	18	4.83 (4.70, 4.97)	1,034 (531; 503)	0.06 (-0.06, 0.18)
Secondary outcome, History, EAL pupils only	33	4.48 (4.31, 4.65)	39	4.53 (4.36, 4.70)	880 (429; 451)	0.04 (-0.10, 0.17)
Secondary outcome, English, EAL pupils only	0	4.80 (4.69, 4.90)	0	4.80 (4.70, 4.90)	1,786 (909; 877)	0.07 (-0.03, 0.17)
Primary outcome Science, EAL+FSM pupils only	13[SDC]	4.39 (4.22, 4.54)	10[SDC]	4.55 (4.36, 4.75)	435[SDC] (223[SDC], 190[SDC])	0.07 (-0.09, 0.22)

Note. As specified in the Statistical Analysis Plan, no p-values are reported. We used bootstrapped CIs to indicate the range of potential effect sizes based on the trial sample. The CIs for all outcomes included '0'. [SDC]Please note that the exact number of missing values for FSM needed to be redacted for SDC; the N in this table is reported with reference to the adjusted N = 435 of pupils with FSM.

Secondary outcome analysis

The analyses for the two secondary outcomes in their respective subsamples also showed no indications of a potential intervention effect. The first of these analyses investigated how effective the 'EAL in the Mainstream Classroom' programme is in improving subject-specific academic attainment in a second GCSE subject (history), in EAL pupils taking history. As presented in Table 12, whilst the intervention group showed again higher scores on average, the adjusted effect size based on the analytic model is small (effect size: 0.04; 95% CI: -0.10 – 0.17) and the CI includes '0'. Therefore, we cannot be confident that there was a difference between the two groups, given the pre-set criteria.

The second of these analyses investigated how effective the 'EAL in the Mainstream Classroom' programme is in improving academic attainment in English (as measured by GCSE English language) when delivered to Key Stage 4 EAL pupils. As presented in Table 12, whilst the intervention group showed again higher scores on average, the adjusted effect size based on the analytic model is small (effect size: 0.07; 95% CI: -0.03 – 0.17) and the CI includes '0'. Therefore, we cannot be confident that there was a difference between the two groups, given the pre-set criteria.

Therefore, the conclusion for both secondary outcome analyses is that the intervention has not shown a potential effect on either outcome and sample. To reiterate from the 'Methods' section, these results cannot be used to gauge the efficacy of the intervention and have an exploratory character.

As detailed in Appendix N, the imbalance analysis detected potential baseline imbalances for variables that were considered in this study. For the secondary analysis investigating GCSE history attainment as an outcome, in addition

to Key Stage 2 SATs reading attainment (KS2_READMRK; which is included in all analyses as a control variable), EAL fluency level and FSM status were found to be imbalanced. The corresponding sensitivity analysis nevertheless concluded that the interpretation remained unchanged: the direction and size of the potential intervention effect (*d* = 0.07) were similar, and the corresponding CI included '0'.

The analysis of imbalance at baseline (see section above) found a potential imbalance for the secondary analysis investigating GCSE English as an outcome for all EAL pupils, with pupils in the control group being assigned to higher proficiency in English bands. The corresponding sensitivity analysis nevertheless concluded that the interpretation remained unchanged: the direction and size of the potential intervention effect (*d* = 0.07) were similar, and the corresponding CI included '0'. An incidental finding of this analysis was that after controlling for Key Stage 2 SATs reading scores (KS2_READMRK) the dichotomised proficiency in English bands predicted GCSE English language scores: A–C on average 0.53 higher (95% CI: 0.35 – 0.71; see Appendix O, Table 8, for more details) GCSE scores than D–E.

Analysis in the presence of non-compliance

Overall, N = 221 teachers were documented in our teacher database. Of these N = 111 teachers, some were in some form involved in the intervention (science or history) and eligible to participate in a workshop. Of these, 58 (52.3%) teachers participated in all three workshop sessions, 26 (23.4%) in two, 15 (13.5%) in one, and 12 (10.8%) in none of the sessions). The compliance analysis focuses only on the primary outcome (science), for which only the science teachers contribute (56 in the intervention condition; 55 in the control). Of these, 31 (55.4%) attended all three workshops; 13 (23.2%) attended two workshops, and 12 (21.4%) attended either one or none of the workshops (exact category N suppressed for SDC). Given that there are three training sessions, compliance was represented on a scale of '0' (no training attended, 0/3 sessions) to '3' (all three sessions attended, 3/3 sessions). We aggregated workshop attendance across teachers for pupils, who experienced the programme with more than one teacher. This situation occurred rarely and only in places where one teacher had already the maximum training, i.e. these contexts remained coded as '3 = maximum dosage'. The Causal Average Complier Effect (CACE) analysis results in an estimate of the relationship between the compliance variable and the outcome (science score, KS4SCI), corrected for potential selection effects of teachers at certain schools being more likely to attend the training (see 'Methods' section; Steele *et al.*, 2007). The analysis details are reported in Appendix O, Table 9. The analysis showed that the compliance indicator was not related to the primary outcome (science score, KS4SCI; effect size across the full range of the compliance indicator: 0.02).

Evaluating the plausibility of the approach, the equations showed that the planned instrumental variables were unlikely to be predictive of compliance (apart from intervention group allocation); and that some of the instrumental variables (Key Stage 2 SATs reading scores, individual and aggregated scores per school) were predictive of the target outcome. These are both undesirable (discussed in detail in Steele *et al.*, 2007). The covariance between the two random effects (i.e. the part of the analytic model that should control for selection effects, if present) was also very small. These results indicate that the combination of variables used in the CACE analysis was empirically not well-suited to predict compliance and to control for selection effects (Steele *et al.*, 2007). But the results also suggest that overall compliance of teachers with training was unlikely to be a major factor correlating with individual progress. (See 'IPE' section for further details relating to training).

Subgroup analyses

The detailed results of the analyses are all available in Appendix O.

The first subgroup analysis assessed the potential impact of the programme on pupils eligible for FSM (EVERFSM_6_P_SPR19). The model was first run only in the sample of EAL pupils taking science classes and eligible for FSM. The resulting effect size estimate was in the same direction as in the primary outcome analysis (effect size 0.07; 95% CI: -0.09 – 0.22), intervention group performing slightly better). The more formal test of this subgroup effect via a cross-level interaction effect (see 'Methods' section for details) found no supporting evidence for such a subgroup effect, i.e. there was no differential (dis-)advantage found for pupils taking science classes depending on whether or not they were eligible for FSM.

The second subgroup analysis evaluated the potential impact of different baseline proficiency in English bands of participating EAL pupils. As described in the 'Methods' section, due to few pupils being evaluated at bands levels A and

B, only a test of lower (A–C) versus higher (D–E) proficiency bands was performed. The analysis found no evidence that pupils of either proficiency band benefited more or less from the intervention.

Finally, the exploratory nature of such subgroup analyses should be considered. The trial was not primarily planned for these analyses and consequently possess neither the sensitivity (statistical power) to detect potentially small effects reliably, nor are they guarded against encountering false-positive findings. The results are only presented to inform potential future developments and evaluations of the programme, or to detect very large effects due to such variables that would require additional investigation and explanation to contextualise the results of this trial.

Additional analyses and robustness checks

Three additional analyses were performed and details for all of these are presented in Appendix O.

First, the impact of 'EAL in the Mainstream Classroom' when pupils experienced the approach from more than one teacher in more than one subject area (i.e. when pupils were taught by 'EAL in the Mainstream Classroom' teachers in both science and history GCSE subjects) was analysed. This analysis was performed in the sample with all EAL pupils and GCSE English language attainment (KS4ENG) as the outcome. In this analysis sample (N = 1,786), 60% (n = 1,071) pupils were taught by at least one science teacher who was part of the project; and 53.3% (n = 952) by at least one history teacher. Across a small and therefore undisclosed number of schools, n = 237 (13.3%) pupils were taught by at least one project teacher in each discipline. We re-ran the same model as for the secondary outcome analysis with GCSE Key Stage 4 English language as the dependent variable and added this exposure score (one versus both subjects) as a pupil-level covariate. Adding this coefficient allowed us to test whether being taught by 'EAL in the Mainstream Classroom' intervention teachers in two disciplines (versus being taught by any number of teachers in only one discipline) had an additional effect over and above programme provision in itself. Whilst we observed a small coefficient in the expected direction (pupils with teachers in both disciplines showed 0.17 higher GCSE English results; 95% CI: -0.01 – 0.35; effect size 0.12; see Appendix O), the coefficient's CI included '0' and we concluded that there is insufficient evidence for a differential effect of being taught by more than one teacher in more than one subject area.

Second, we tested whether an EAL student's choice regarding combined science/three separate science GCSEs introduced heterogeneity in the programme's effect due to possibly different profiles of pupils taking these options. We re-ran the same model as for the primary outcome analysis including an interaction effect for combined science/three separate science GCSEs. In the primary outcome data set N = 402 (37.5%) pupils had selected three separate sciences according to their NPD result scores and N = 669 (62.5%) combined science. Descriptively, these two student groups differed in their outcomes, with pupils who had selected three separate sciences achieving an average of 5.85 (SD = 1.59) and pupils with combined science GCSEs achieving an average of 4.18 (SD = 1.78; see Table 11). Whilst our analysis showed also after controlling for school clustering and Key Stage 2 reading scores that pupils who selected three sciences had on average a 1.13 (95% CI: 0.85 – 1.42; see Appendix O, Table 11, for more details) higher GCSE science score than those who did not, the CI for the cross-level interaction effect with treatment included '0' (95% CI: -0.63 – 0.24; see Appendix O, Table 11, for more details), i.e. pupils in either science GCSE were unlikely to differentially benefit from the intervention. The conclusion for the effect of the intervention as evaluated in the primary analysis therefore, remained unchanged.

Third, we tested the impact of the 'EAL in the Mainstream Classroom' programme on non-EAL pupils within the same classrooms. We: (a) used all available student data, i.e. EAL and non-EAL pupils; and (b) added a cross-level interaction effect between EAL/non-EAL and the condition. The analysis found no evidence that either EAL or non-EAL pupils benefited more or less from the intervention (maximal effect size difference shows that EAL pupils in the control group did on average better than non-EAL pupils in the intervention group, with an effect size of 0.05). An incidental finding of this analysis is that EAL pupils performed better than FLE pupils after controlling for Key Stage 2 SATs reading attainment (KS2_READMRK) (regardless of intervention allocation; on average 0.46 higher GCSE English language scores).

IPE results

The IPE was designed to address the logic model and the two main IPE research questions, namely:

1. The extent to which the programme was implemented with fidelity, and what modifications, if any, were adopted?
2. What is the impact of 'EAL in the Mainstream Classroom' programme on classroom experience?

Using the methods described above and in Table 6, this section discusses:

- the reach of the programme, including recruitment to the evaluation;
- business as usual i.e. the control condition, and baseline characteristics in both control and intervention schools;
- experience of training including attendance at training and completion of intersessional tasks;
- teachers' learning through CPD provided by the programme;
- fidelity in implementation, adaptations, and what the intervention 'looks like' in the classroom;
- perceived student outcomes (as perceived by intervention teachers); and
- dosage, in terms of the amount of time pupils were taught by participating teachers.

Reach

Recruited schools had a high proportion of EAL pupils and additional training to support EAL pupils was an identified need by participating schools.

- Overall, 40% of participating science teachers were in a leadership role.
- Nearly three-quarters of Year 10 science classes were studying for Combined Science GCSE Awards and 26% for the three separate awards of GCSEs in biology, chemistry, and physics.
- 30% of Year 10 science classes were taught all three science subjects by a participating teacher and 60% were taught one science subject only (i.e. biology, chemistry, or physics).
- 27% of classes were taught the full science curriculum (i.e. all three sciences) by a participating teacher and approximately 47%, a third of the science curriculum (i.e. one of either biology, chemistry, or physics).
- Where classes were set by ability Year 10 science classes taught by participating teachers were fairly evenly spread across the different prior attainment levels.

Reach is defined for the purposes of this report in three different ways: 1) the schools recruited; 2) participating teachers; and pupils in the participating classes. This includes both those allocated to receive the intervention and those allocated to the control condition to assess comparability. In addition, information is provided surrounding the amount of curriculum time participating classes were taught by participating teachers. This is particularly important for those in intervention classes as, given that teachers' learning from the CPD was designed to be embedded within their lesson delivery, it indicates the extent to which pupils could be exposed to elements of the 'EAL in the Mainstream Classroom' content.

School recruitment

As indicated in the sections above on 'Trial design' and 'Participant flow and attrition', recruitment to the trial was lower than anticipated. Recruitment was led by the Delivery Centres with support from Challenge Partners (with the exception of the late recruited Delivery Centre as described in the 'Recruitment' section above). Delivery Centres chose their own organisational frameworks for delivering this remit, which varied quite significantly. At one extreme, the Head of English (who was also the Assistant Head with special responsibility for teaching, learning, and CPD) undertook the organisation of the whole programme, as well as co-delivering two of the three workshops, in addition to their ongoing full-time commitments. At the other extreme, one Delivery Centre wrote the administrative duties of delivering the EAL training into the job description of a new member of the general office staff, so that the facilitators could concentrate solely on the delivery of the training workshops. A variety of largely pragmatic, less clearly defined, arrangements were adopted by the other Delivery Centres. In terms of recruitment however, all Delivery Centres were provided with lists from Challenge Partners of possible schools in their area (i.e. schools with high proportions of EAL pupils) and interviewees reported receiving good support in general from the delivery team; '*they [Challenge Partners] were proactive*'. Delivery Centre members also discussed using their own networks and contacts to recruit schools, for example:

> *I've got a lot of connections, I'm in several networks…and also I've taught in a few schools before, I provided some training for other schools, so I was able to recruit quite a few other schools to come on board myself, but based on personal contacts.* (Delivery Centre facilitator interview)

Recruitment was, however, a time-consuming process:

> *I could just sit in there for about an hour and just dial up, and email, and chase, and harry [sic], and harass I think to a degree, but yes, it worked quite well.* (Delivery Centre facilitator interview)

Delivery Centres were requested to recruit ten secondary schools each, targeting schools:

- with at least 15% or more learners for whom English was not the first language across the school;
- willing to release at least one teacher in science and one teacher in history who in 2017/2018 would be teaching Year 10 GCSE classes containing at least 14 EAL pupils expected to enrol on a GCSE history programme and 14 EAL pupils taking a GCSE science programme;
- be within reasonable travelling distance of the Delivery Centre; and
- not be implementing the programme or intend to acquire the programme until after summer 2019 if allocated to the control condition (Challenge Partners, 2017b).

As intended, the schools recruited did have a high proportion of EAL pupils, with an average of 44% of EAL pupils across schools (see Table 10). Teacher interviews indicated that additional training to support EAL pupils was an identified need within their school, whether this be an existing high EAL population or an increasing EAL population due to changing local demographics:

> *The increased population of EAL students in the school…so coming in from other cultures. It's something that we're aware of and it's one of the PP [Pupil Premium] or the disadvantaged indicators, so I think that's where the school's looking at trying to develop.* (Teacher interview)

Schools aiming to improve attainment levels for EAL pupils was frequently mentioned, and less often, within this, the subject specialisms for the evaluation (science and history):

> *The English we've always known as being a problem on exams, for example, the reading age of most science exams has been 16 plus which can be challenging for a normal native speaker and obviously for our EAL cohort that can be, you know, a real big challenge.* (Teacher interview)

Participating teachers

At baseline, there were 116 (60 intervention, 56 control) science teachers participating in the evaluation. Of these 116, the survey was completed by 106 science teachers at T1 (60 intervention, 46 control; 92% completion rate), 73 science teachers at T2 (42 intervention, 31 control; 63% completion rate) and 38 intervention teachers at T3 (63% completion rate). It is notable that completion rates were lower for control teachers at each timepoint they were requested to complete the survey (T1 and T2; 82% and 55%, respectively). This can be explained by the fact that a small number of intervention teachers completed the survey after randomisation and control schools had no incentive to inform the research team if a change of teacher had occurred for the participating evaluation classes.

As indicated in the logic model, participating teachers were intended to be drawn from science and history departments in participating schools and to be 'emerging leaders', thereby having a greater capacity to embed the programme within schools more widely. In the baseline survey, participating science teachers were asked about their roles within the school. As seen in Table 13, 40% (or 42 out of 106 teachers) indicated that they held a leadership role in their school (21 out of 60 intervention teachers or 35% and 21 out of 46 control group teachers or 46%).

Table 13: Leadership roles of participating teachers*

	Intervention n (%)	Control n (%)	Total n (%)
Leadership role	21 (35)	21 (46)	42 (40)
Emerging[1]	10 (17)	14 (30)	24 (23)
Science-related *(e.g. Second in Science)*	8 (13)	10 (22)	18 (17)
School-related *(e.g. Head of Year)*	2 (8)	4 (9)	6 (6)
Established	9 (13)	5 (11)	14 (13)
Science-related *(e.g. Head of Department)*	6 (10)	4 (9)	10 (9)
School-related *(e.g. Assistant Headteacher)*	3 (5)	1 (2)	3 (3)
Role not stated	2 (3)	2 (4)	4 (3)

**N = 106 (intervention = 60, control = 46).*
Percentages may not add up to 100 due to rounding.
[1] Emerging leaders was taken as meaning those who had recently taken on school leadership roles or were expected to do so in the near future.

Teachers were also asked in the baseline survey (T1) how long they had been in the teaching profession and the number of years working in their current school. Their responses are given in Table 14. As can be seen, over a third (36%) reported that they had been in the teaching profession between two and five years (38 teachers out of 106) indicating that they were establishing themselves in the profession[27] and over half had been in their current school two–five years (53% or 56 teachers out of 106) and therefore potentially establishing themselves within the school. Taken together, Tables 13 and 14 show that although schools were requested to nominate participating teachers prior to randomisation, there is an imbalance between teachers in the intervention and control conditions, with those teachers allocated to control being more likely to report holding a leadership role and to have been in the profession longer (six years and more), which could have implications for any findings.

Table 14: Number of years teaching and number of years in current school*

Years	Number of years teaching			Number of years in current school		
	Intervention n (%)	Control n (%)	Total n (%)	Intervention n (%)	Control n (%)	Total n (%)
0–1 year	4 (7)	4 (9)	8 (8)	17 (28)	10 (22)	27 (26)
2–-5 years	26 (43)	12 (26)	38 (36)	31 (52)	25 (54)	56 (53)
6–10 years	10 (17)	9 (20)	19 (18)	6 (10)	2 (4)	8 (7)
11–15 years	10 (17)	9 (20)	19 (18)	4 (7)	4 (9)	8 (8)
16+ years	59 (17)	12 (26)	21 (20)	2 (3)	5 (11)	7 (7)

**N = 106 (intervention = 60, control = 46).*
Percentages may not add up to 100 due to rounding.

Teachers' reported number of years teaching and number of years in their current school does not, however, given an accurate picture of 'emerging leadership', nor does existing role. Intervention teachers in the interviews were asked

[27] There are a number of debates surrounding the definition of an 'experienced teacher'. However, many fall somewhere between the third and fifth years of teaching, hence the definition here of two–five years as an 'emerging leader', i.e the teacher is experienced and potentially ready to take on new/extended responsibilities within a school (Graham *et al.,* 2020).

about how they came to participate in the evaluation. This data was subsequently analysed to establish if career progression or leadership skills was mentioned (unprompted). Only two respondents mentioned either of these facets:

> *I'm also head of Year 10 [the year group of participating pupils] so I thought to myself why not just bring a new initiative into the school and see if we can find further ways that [improve] attainment of some of the pupils in our year group.* (Teacher interview)

Others saw the programme as a potential to improve their teaching practice and were willing to: '*go and see what it was about*'. Only one interviewee stated that they were on the programme because: '*I was told, basically*' although this respondent also felt that they were an appropriate selection for the programme:

> *…we're all experienced teachers who would be confident in trying some of these methodologies in our classes and we'll reflect on them.* (Teacher interview)

Participating Year 10 classes

In the baseline survey (T1), science teachers were asked how many Year 10 classes they taught that were participating in the trial, what science subject(s) they taught their Year 10 classes, what proportion of the curriculum they taught, the qualification the pupils were aiming for, and the ability levels of those classes.

Number of classes

Teachers reported teaching up to four Year 10 participating classes: 63% of teachers (67 out of 106) taught only one Year 10 science class, 26% (28 out of 106) taught two Year 10 science classes, 9% (10 out of 106) taught three classes and 1% (1 out of 106) taught four classes. Altogether this represents 157 classes taught across 106 teachers (Table 15). Again, there is some disparity between control and intervention schools with intervention teachers more frequently mentioning that they only taught one Year 10 class in the evaluation. It should also be noted that the survey returns on the number of participating classes taught differ markedly from the class UPN data provided by the schools prior to randomisation. Whilst the survey data from 106 teachers records a total of 157 classes, the UPN data for the 113 science teachers who participated in the trial records a total of 132 classes, with only one teacher teaching more than two classes and only 13 teachers teaching two classes. The survey figures, therefore, should be read with some caution as it is likely that a number of teachers neglected to distinguish between the number of classes they were teaching as a whole and the number that were actually participating in the trial. These figures do, however, perhaps more accurately reflect the number of Year 10 classes who experienced some elements of the programme (by virtue of being taught by the same teacher). Consequently, the figures reported below (Tables 15–19) are also more likely to reflect the extent of Year 10 science teaching provided across the evaluation cohort and, given the nature of the programme, there is the potential for teaching based on strategies learned across this cohort (i.e. the reach was potentially larger than the number of intervention pupils included in the evaluation).

Table 15: Number of classes taught, as reported by teachers, T1*

	Intervention n (%)	Control n (%)	Total n (%)
One class only	42 (70)	25 (54)	67 (63)
Two classes	13 (22)	15 (33)	28 (26)
Three classes	4 (7)	6 (13)	10 (9)
Four classes	1 (2)	0 (0)	1 (1)

**N = 106 (60 intervention teachers, 46 control teachers).*
Percentages may not add up to 100 due to rounding.

Target qualification

Of the 156 classes for whom the target qualification was reported, approximately three-quarters of both control and intervention classes were studying for the combined science GCSE. These were 55 classes out of 73 or 75% in the control group and 59 classes out of 83 or 71% in the intervention group. The remaining classes, with one exception

which was studying for an entry level certificate in science, were studying the three separate science subjects (see Table 16).[28]

Table 16: Target qualification for Year 10 science classes*

	Intervention n (%)	Control n (%)	Total n (%)
Combined science**	59 (71)	55 (75)	114 (73)
Three separate sciences***	24 (29)	17 (23)	41 (26)
Other****	0 (0)	1 (1)	1 (1)

** N = 156 classes (83 intervention classes, 73 control classes).*
Percentages may not add up to 100 due to rounding.
*** Combined science awards two GCSEs across all three science subjects.*
**** Three separate sciences refer to pupils undertaking a GCSE in each of the individual science subjects of biology, chemistry, and physics, which is the alternative to combined science.*
***** Entry Level Certificate in Science.*

Subject(s) taught

60% of classes (90 out of the 149 classes reported on) were taught one science subject by participating teachers, whether that be biology, chemistry, or physics, although for control classes this was higher at 67% (or 47 classes out of 70) and for intervention classes lower, at 54% (or 43 classes out of 79). In approximately a third of classes (45 classes out of 149 or 30%), teachers taking part in the evaluation taught all three science subjects, with intervention teachers more likely to report teaching all three sciences (38% or 30 classes out of 79) compared to those in the control schools (21% or 15 classes out of 70). The remainder (9% or 14 classes out of 149) were taught two out of the three science subjects by the evaluation teacher (11% or 8 out of 70 control classes and 8% or 6 out of 79 intervention classes) (see Table 17).

Table 17: Subject(s) taught to Year 10 science classes*

Science subject taught	Intervention n (%)	Control n (%)	Total n (%)
Biology, Chemistry, Physics	30 (38)	15 (21)	45 (30)
Biology only	16 (20)	12 (17)	28 (19)
Chemistry only	16 (20)	21 (30)	37 (25)
Physics only	11 (14)	14 (20)	25 (17)
Biology, Chemistry	3 (4)	4 (6)	7 (5)
Biology, Physics	1 (1)	2 (3)	3 (2)
Chemistry, Physics	2 (2)	2 (3)	4 (3)

** N = 149 classes (79 intervention classes, 70 control classes).*
Percentages may not add up to 100 due to rounding.

[28] This class was not, in fact, a class participating in the evaluation. As indicated above, there appears to have been some uncertainty among teachers which Year 10 science classes they taught were included in the evaluation. This may be because the pupil data was usually organised and provided to the evaluation team by one member of the school staff rather than by individual participating teachers.

Ability level of classes

Teachers also provided information relating to the ability levels of pupils for 152 classes.[29] Thirty-six classes out of 152 (or 24%) were regarded as being set as low ability (with a slightly higher proportion of control classes being set in this category compared to intervention classes (19 classes out of 73 or 26% of control classes and 17 classes out of 79 or 22% of intervention classes). Overall, however, Year 10 science classes taught by participating teachers were fairly evenly spread across the ability levels (from high ability [19% of classes or 29 classes out of 152] to low ability classes [24% or 36 classes out of 152]), with a further 9% of classes (13 classes out of 152) being of mixed ability (Table 18). Perhaps unsurprisingly, those in the higher ability groups ('High', 'Medium High') were also most likely to be in the classes taking the three separate subject GCSE sciences.

Table 18: Ability levels of participating classes*

	Class ability level*		
	Intervention n (%)	Control n (%)	Total n (%)
Set by ability			
High	15 (19)	14 (19)	29 (19)
Medium/high	14 (18)	7 (19)	21 (14)
Medium	14 (18)	11 (15)	25 (16)
Medium/low	14 (18)	14 (19)	28 (18)
Low	17 (22)	19 (26)	36 (24)
Mixed ability	5 (6)	8 (11)	13 (9)

** N = 152 classes (79 intervention classes, 73 control classes).*
Percentages may not add up to 100 due to rounding.

Proportion of curriculum taught

Approximately a quarter (27% or 41) of the 155 classes for whom data was obtained were reported to receive their teaching for the full science curriculum by a teacher participating in the evaluation (Table 19). This roughly approximates to the same number of classes reported as being taught all three science subjects by a participating teacher (Table 17). Seventy-two classes out of 155 classes reported on (or 47%) were taught between 20% and 39% of their science time by a participating teacher. This represents a third of the science curriculum and, again, corresponds closely with the number of classes reported as receiving teaching in only one science subject by the participating teacher. Discrepancies are, in part due to the different number of classes reported on for subject(s) taught and proportion of time spent teaching classes (149 and 155 classes, respectively). When the proportion of curriculum taught is taken together with the science subject(s) taught (see Table 17 above) it can be seen that, for those pupils in the intervention group, they received either all their science teaching from an 'EAL in the Mainstream Classroom' teacher or all science content in one subject (biology, chemistry, or physics) from an 'EAL in the Mainstream Classroom' teacher.

Table 19: Proportion of Year 10 science curriculum taught by teachers taking part in the evaluation to Year 10 science classes, T1*

	Intervention n (%)	Control n (%)	Total n (%)
100%	22 (27)	19 (26)	41 (27)
80–99%	4 (5)	1 (1)	5 (3)
60–79%	8 (10)	4 (5)	12 (8)
40–59%	18 (22)	8 (11)	26 (17)
20–39%	30 (37)	42 (57)	72 (47)
0–19%	1 (1)	0 (0.0)	1 (1)

** N = 155 classes (81 intervention classes, 74 control classes).*
Percentages may not add up to 100 due to rounding.

[29] The researchers acknowledge that 'set by prior attainment' would have been more appropriate terminology although at the time of the evaluation 'set by ability' was more widely used in schools.

Usual practice

At T1 (baseline), nearly two-thirds of respondents felt that their school supported EAL learners either 'well' or 'quite well'. The most frequently reported provisions within schools to support EAL pupils were 'the use of a buddy or peer support system for EAL pupils', 'one to one specialist support for new EAL pupils' and 'modelling by peers and/or teachers' (43 schools out of 65 or 66%). However, in some cases, there appeared to be a lack of awareness of in-school support for EAL pupils.

- In contrast, less than half of T1 survey respondents indicated that they felt 'very well' or 'quite well able' to support the EAL pupils in their classroom and less than a third felt confident in supporting EAL pupils.
- Similarly, nearly a half of teachers felt that their EAL pupils understood the importance of subject-specific language 'very' or 'quite' well and nearly a third of teachers reported that EAL pupils in their science classes were confident using subject-specific language.
- Teachers were more positive about their EFL pupils: Over three-quarters of respondents (reported that their EFL pupils understood the importance of subject-specific language 'very' or 'quite' well and over two-thirds reported that the EFL students in their science classes were confident using subject-specific language.

This section draws on survey data to describe usual practice in the participating schools, primarily at baseline (T1), for both control and intervention schools. It also describes the wider support within schools, outside of the programme, as measured in the end of the evaluation survey (T2). It does so to assess any additional changes, outside of 'EAL in the Mainstream Classroom' programme, in provision for EAL pupils in participating schools overall in both control and intervention schools (perhaps as a result of participating in the programme), as well as approaches and understanding of subject-specific language teaching practices. In addition, this section explores the way in which usual practice is experienced including student engagement, understanding of, and perceived importance of, subject-specific language. This provides a baseline by which the intervention can be measured against perceived outcomes of the programme in latter sections of this report as well as identifying any possible differences at baseline between control and intervention schools.

The survey was completed by 106 of the 116 science teachers taking part in the evaluation at baseline (T1). For intervention teachers the number of survey returns and completion rates were as follows: at T1, 60 out of 60 intervention teachers, 100% completion; at T2, 42 out of 60 teachers, 70%; and at T3, 38 out of 60 teachers, 63%.[30] The corresponding returns and completion rates for those science teachers in the control group were: at T1, 46 out of 56 teachers, 82%; and at T2, 31 out of 56 teachers, 55%.

Support for EAL learners

In T1, survey teachers were asked how well they felt EAL learners were supported in their school. Nearly two-thirds of respondents (65% or 69 teachers out of 102) felt that their school supported EAL learners either 'very well' or 'quite well' although control group teachers were more likely to rate this support 'very well' than intervention teachers (16% or 7 teachers out of 45 and 9% or 5 teachers out of 57, respectively) (Table 20).

[30] Whilst 60 teachers later allocated to the intervention group were participating in the evaluation at baseline, four subsequently withdrew from the evaluation after randomisation, only one of whom was replaced by another science teacher within the school. Their data was, however, retained for analysis purposes.

Table 20: Teachers' perceptions of their schools overall support for EAL pupils, baseline survey (T1)*

	Allocation		Total n (%)
	Intervention n (%)	Control n (%)	
Very well	5 (9)	7 (16)	12 (12)
Quite well	33 (58)	24 (53)	57 (56)
Neutral	12 (21)	9 (20)	21 (21)
Not very well	7 (12)	5 (11)	12 (12)
Not very well at all	0 (0)	0 (0)	0 (0)

** N = 102 (57 intervention teachers, 45 control group teachers).*
Percentages may not add up to 100 due to rounding.

Science teachers were then asked a series of questions about what this support for EAL pupils looked like (i.e. what provisions were made). When the responses are collated by school it can be seen that the most extensive provision involved the use of a buddy or peer support system for EAL pupils (48 schools out of 65 or 74%), followed by one to one specialist support for new EAL pupils (48 schools out of 65 or 74%) and modelling by peers and/or teachers (44 schools out of 65 or 68%) (Table 21). It can also be seen that there were some differences in provision for EAL pupils within schools between the control and intervention groups. For example, 80% of control schools (24 schools out of 30) were reported to have a buddy or peer support system in place for pupils compared to 69% of intervention schools (24 schools out of 35). In contrast, 49% of intervention schools (17 out of 35) were reported to have wall displays and classroom resources to support specifically EAL children compared to only 30% of control schools (9 schools out of 30). Whilst two schools were reported to have no provisions in place, those who did reported having between one and ten of the listed provisions (with an average of 5) in both control and intervention.

Table 21: Provision in schools for EAL pupils, T1*

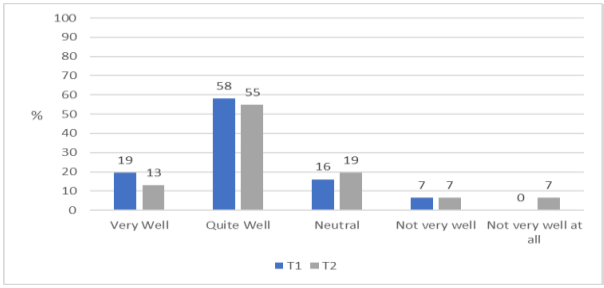
	Allocation**		
	Intervention n (%)	Control n (%)	Total n (%)
Use of buddy or peer support system for EAL pupils	24 (69)	24 (80)	48 (74)
One to one specialist support for new EAL pupils	27 (77)	21 (70)	48 (74)
Modelling by peers and/or teachers	22 (63)	22 (73)	44 (68)
Specific differentiation for EAL pupils in the classroom	20 (57)	18 (60)	38 (59)
Language specialist teaching assistants in the classroom	16 (46)	14 (47)	30 (46)
Specialised visual aids and tools for EAL learners	14 (40)	15 (50)	29 (45)
Wall displays and classroom resources to support specifically EAL children	17 (49)	9 (30)	26 (40)
Interactive translation tools and technology	14 (40)	12 (40)	26 (40)
Use of actions, signing, or miming	13 (37)	5 (17)	18 (28)
Talk frames and oral rehearsal	8 (23)	8 (27)	16 (25)
Specific resources***	5 (14)	4 (13)	9 (14)
None	1 (3)	1 (3)	2 (3)

** More than one response could be given.*
*** N = 65 schools (intervention = 35 schools; control = 30 schools).*
Percentages may not add up to 100 due to rounding.
**** For example, bilingual dictionaries, non-specialist teaching assistants.*

In addition to the provisions listed in the survey above, five teachers mentioned the use of bilingual dictionaries as specific resources available to EAL pupils, two teachers mentioned the availability of non-specialist teaching assistants, and four teachers mentioned pupils being specifically withdrawn from mainstream lessons for additional language learning (either all lessons, or some lessons). Two teachers, from separate schools, indicated that there was no provision for EAL pupils in their school (although for one of these teachers a colleague within the same school did list some provision). It should be noted, however, that teachers from the same schools did sometimes provide different pictures of support, indicating that, in some cases, there may have been a lack of awareness of exactly what support was available for EAL pupils.

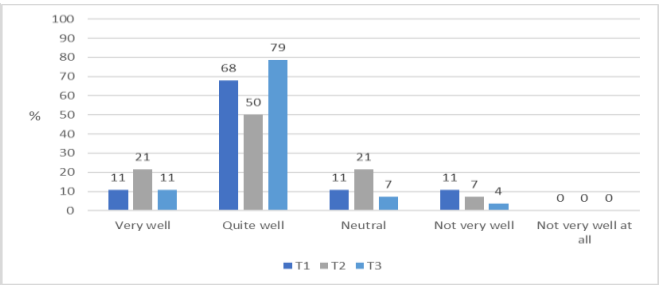
At T2 (the end of the first, and primary, year of the evaluation), teachers were again asked about how well they felt their school supported EAL pupils and the forms of provision available. Data relating to the teachers at the school level by allocation is presented in order to assess if any change had occurred over time. Given the potential for individual change responses were only included for those participants who answered this question at all timepoints.

Figure 5: Control teachers' perceptions of their school's overall support for EAL learners at T1 and T2, teachers who completed at both timepoints only*



** N = 31.*
Percentages may not add up to 100 due to rounding.

Figure 6: Intervention teachers' perceptions of their school's overall support for EAL learners at T1, T2 and T3, teachers who completed at all three timepoints only*



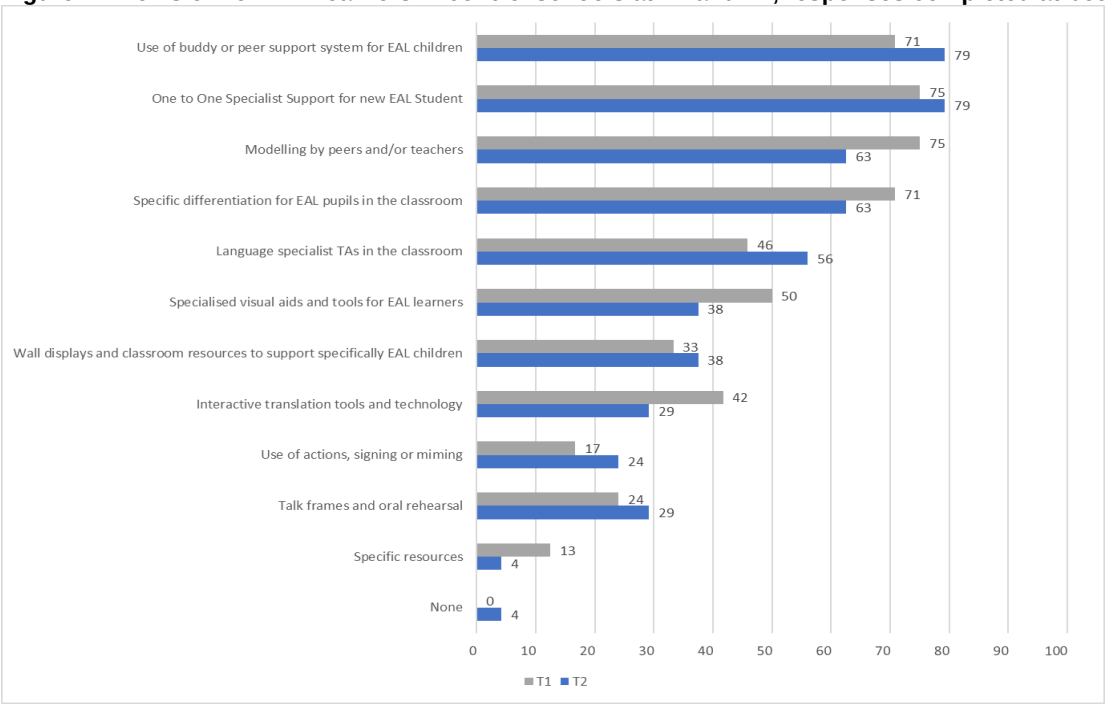
** N = 28.*
Percentages may not add up to 100 due to rounding.

As seen in Figure 5, although numbers are small, teachers in control schools showed a downward trend in perceptions of their school's overall support over time (i.e. between T1 and T2). For example, 19% (or 6 teachers out of 31) indicated that they felt that their school supported EAL learners 'very well' at TI but only 13% (or 4 teachers out of 31) did so at T2. In addition, at T1 no control teacher felt EAL learners were supported 'not very well at all' but at T2 this had risen to 7% (or 2 teachers out of 31). Figure 6 shows a somewhat different trend for intervention schools. The number of teachers feeling that their school supported EAL learners 'very well' doubled between T1 and T2 (from 11% or 3 teachers to 21% or 6 teachers out of 28), although the numbers who felt their school supported EAL learners 'quite well' fell by more than this gain (from 68% or 19 teachers out of 28 at T1 to 50% or 14 teachers out of 28 at T2). There was also a dilution of implementation of the training over time (by T3 numbers returned to 11% or 3 teachers out of 28 rating this support 'very well' overall) although by T3 teachers were in general feeling more positive than they had done in either T1 and T2 with a shift away from ratings of 'not very well' towards 'quite well'. Overall, however, there was little change in attitudes or provision, any change was of a small magnitude and, given the small number of teacher responses any conclusions should be approached with caution.

Given the variability in awareness of school provisions identified at T1, responses at T1 and T2 (and at T3 for intervention participants) were analysed to only include those respondents who completed this question at each timepoint. Figure 7 shows a somewhat mixed picture of provision in control schools for EAL learners. Teachers demonstrated an increase in, or increased awareness of, some aspects of provisions within schools for EAL learners and potentially a loss of some provision, or an awareness that it was not actually taking place as previously thought. In contrast, awareness of, or actual, provision tended to increase between TI and T2 for intervention schools, with language specialist teaching assistants being the only exception (Figure 8). By T3 the picture is more mixed although, again, there was an increase in reported in-school provision for many aspects listed, or levels remained the same. However, levels of awareness of some forms of support rose at T2 to only fall again at T3. These included: a buddy or peer support system; wall displays and classroom resources; interactive translation tools and technology; talk frames and oral rehearsal; and specific resources.

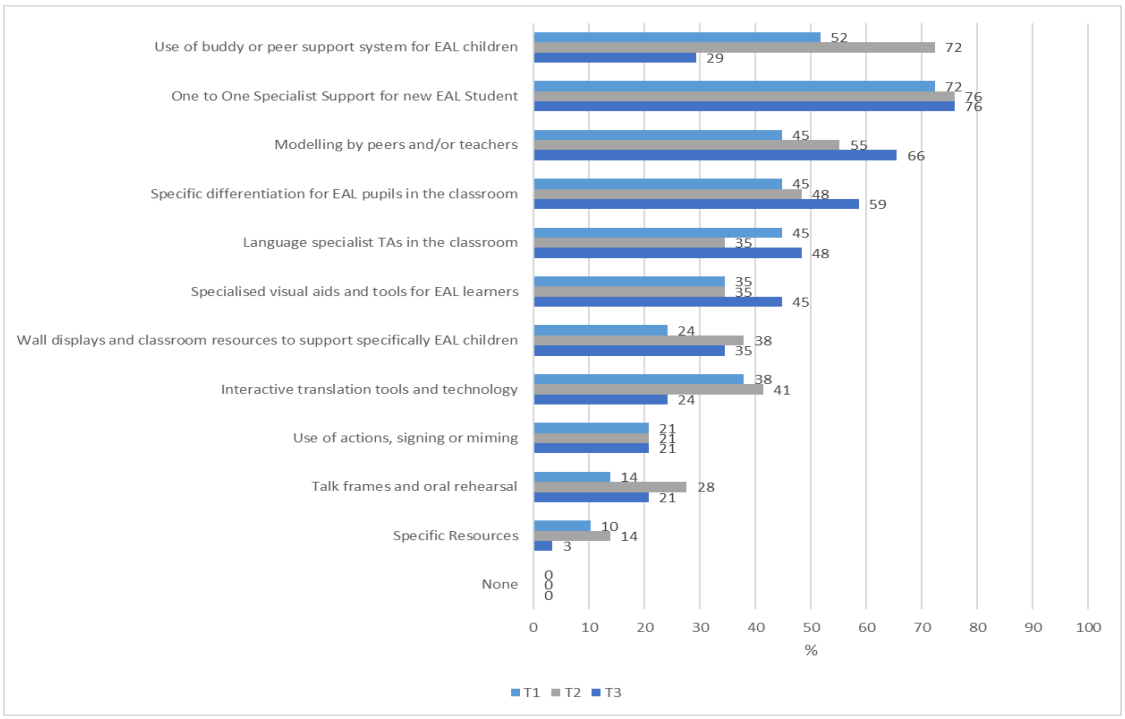
A final form of provision, not asked about in the survey and not provided in the open responses but mentioned in the teacher interviews was school provision for liaising with parents. However, given that this was provided at a school level and seen as part of the 'specialist support' provided by schools, teachers may not have been as acutely aware of this aspect of provision, or have seen it as of direct relevance to their classroom practice.

Figure 7: Provision for EAL learners in control schools at T1 and T2, responses completed at both timepoints only*



** N = 24 control schools. Responses collated from 31 teachers. More than one response could be given. Percentages may not add up to 100 due to rounding.*

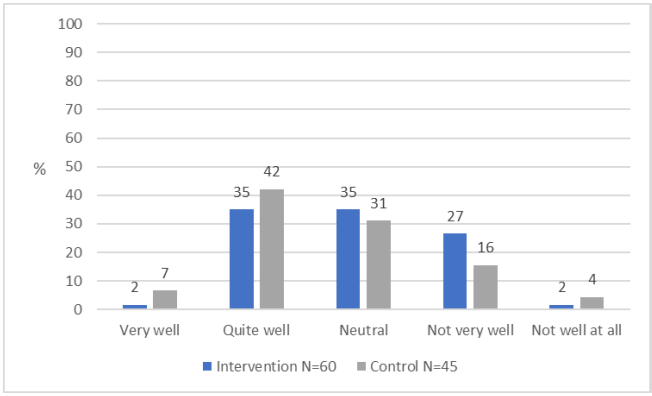
Figure 8: Provision for EAL learners in intervention schools at T1, T2, and T3, responses completed at all timepoints only*



** N = 28 intervention schools. Responses collated from 34 teachers. More than one response could be given. Percentages may not add up to 100 due to rounding.*

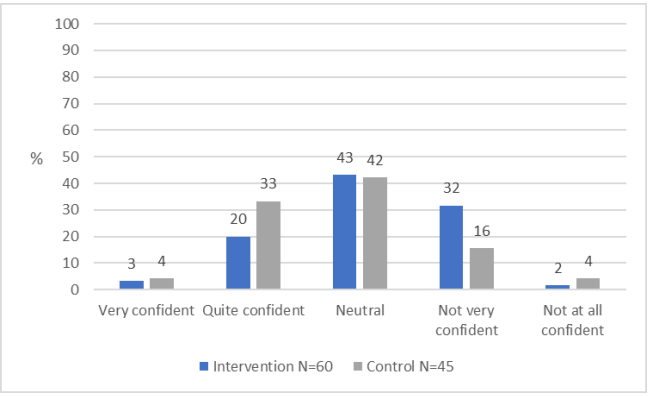
Teachers were also asked in the T1 survey about their own support for EAL learners in the Year 10 classes. Overall, 42% of respondents (or 44 teachers out of 105) indicated that they felt 'very well' or 'quite well able' to support the EAL pupils in their classroom compared to 25% (or 26 teachers) who reported that they felt 'not very well' or 'not well at all' able to support their EAL learners. In contrast, 30% of teachers (or 31 out of 105 teachers) felt 'very confident' or 'quite confident' compared to 25% (26 teachers out of 105) who reported feeling 'not very confident' or 'not at all confident' in supporting EAL pupils. Figures 9 and 10 provide the responses by group allocation and suggest that control teachers overall felt better able to support the EAL pupils in their classrooms and felt more confident to do so than their intervention peers. A comparison of teachers' responses to this and subsequent questions at T1 and T2 is provided in the sections on 'Teacher learning' and 'Perceived impact' below.

Figure 9: Overall, how well do you feel you are able to support the EAL pupils in your classroom in their learning? T1



Percentages may not add up to 100 due to rounding.

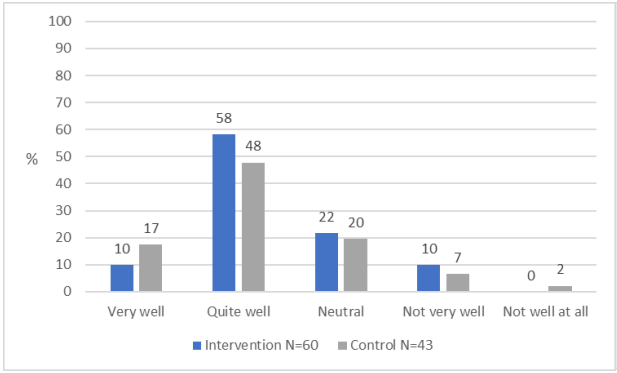
Figure 10: How confident are you in supporting the EAL children in your classroom in their learning? T1



Percentages may not add up to 100 due to rounding.

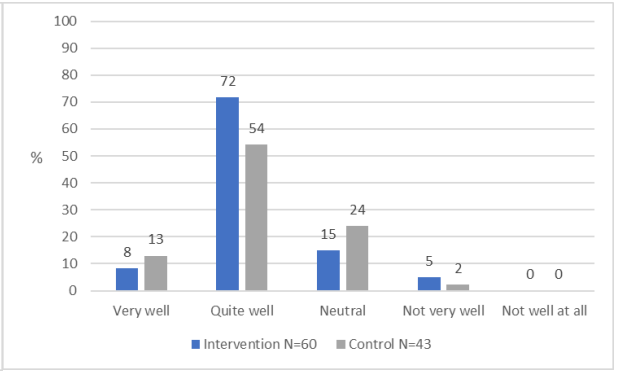
Finally, teachers were asked about their pupils' engagement in their lessons, for both EAL pupils and FLE pupils. The responses by allocation are presented in Figures 11 and 12. Overall, 69% (or 71 teachers out of the 103 who responded to this question) reported that their EAL pupils were 'very' or 'quite' well engaged during their lessons as a whole. A slightly higher proportion of teachers (75% or 77 teachers out of 103) reported that their FLE pupils were 'very' or 'quite' well engaged. Higher proportions of teachers also rated their EAL pupils as 'not very well' or 'not well at all' engaged in their lessons compared to their FLE peers (9% or 10 respondents out of 103 and 4% or 4 respondents out of 103, respectively).

Figure 11: How well do you feel your EAL pupils are engaged during your lessons on the whole? T1



Percentages may not add up to 100 due to rounding.

Figure 12: How well do you feel your non-EAL pupils are engaged during your lessons on the whole? T1



Percentages may not add up to 100 due to rounding.

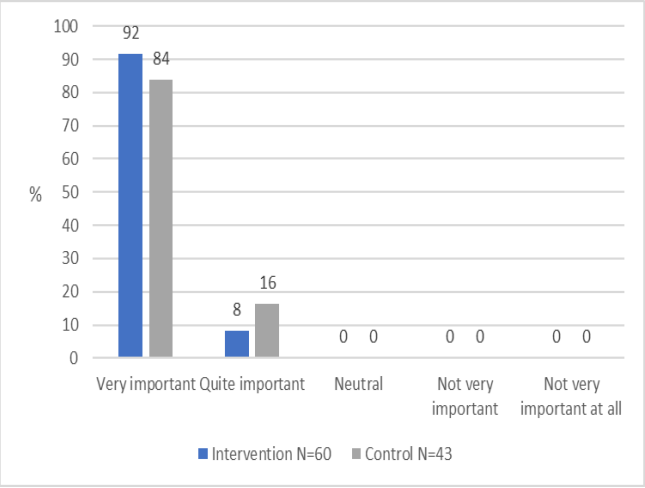
Subject-specific language

Given that the aim of the programme was to increase teachers understanding of the linguistic demands and academic characteristics of their subject and apply that understanding to their teaching it was considered important to assess this at the teacher and pupil levels within schools at baseline (and subsequently ask the same questions at the end of the evaluation to measure change; the findings of this are reported in later sections on fidelity and perceived outcomes, below) as differences at baseline may have had subsequent effects for the overall outcomes of the study. Similarly, given that the CPD was designed to enable teachers to embed subject-specific language teaching within their lessons it was deemed important to understand the extent to which this already took place or if it was dealt with separately within lessons. Consequently, teachers were asked a series of questions relating to their subject-specific language teaching.

At T1, all respondents (100% or 103 teachers) felt that subject-specific language was important to the science subject they taught their Year 10 classes (88% or 91 teachers out of 103 rating it as 'very important' and 12% or 12 teachers out of 103 rating it as 'quite important'). The majority of teachers (87% or 90 of the 103 respondents) felt confident (43% or 44 teachers out of 103 'very' and 45% or 46 teachers out of 103 'quite') in teaching and using subject-specific language in their Year 10 science classes. In addition, when asked about their current approach to subject-specific language teaching in their Year 10 science classes, in the T1 survey, 83% of the 103 teachers (85 teachers) who responded to this question stated that they frequently referred to it in their everyday lesson vocabulary, 8% (8 teachers out of 103) stated that it was taught as a separate element in a lesson and then referred back to and 5% (5 teachers out of 103) stated that it was present in teacher assessments and formative assessments only. Four per cent of teachers (4 out of 103) stated that it was not included in their teaching at all and one teacher stated, in the free-text box under 'Other', that '*using key words which are subject-specific is a regular task*' but provided no further detail as to what this involved.

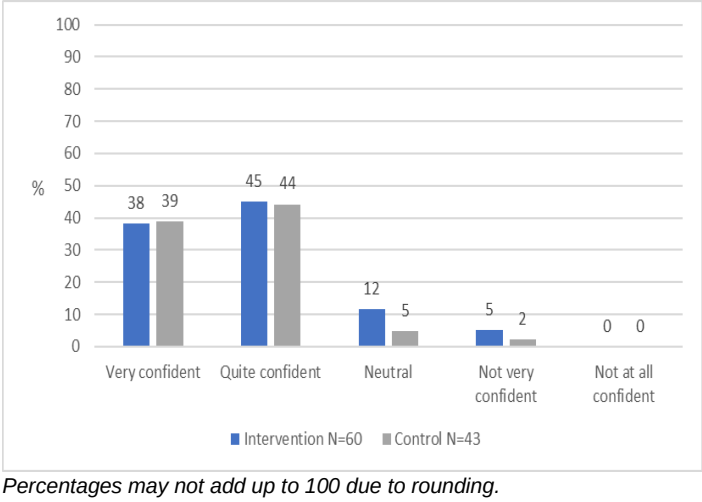
The responses to these questions by allocation are presented in Figures 13 and 14 and Table 22. They show similar patterns of responses across intervention and control groups with the only notable differences being that intervention teachers were slightly more likely to rate subject-specific language as 'very important' in the science subject they taught at T1 compared to control teachers (91% or 55 teachers out of 60 and 83.7% or 36 teachers out of 43, respectively) and a slightly higher proportion of control teachers than those in the intervention group indicated that they *'frequently referred to subject-specific language in everyday lesson vocabulary'* (84% or 36 teachers out of 43 and 82% or 49 teachers out of 60, respectively).

Figure 13: To what extent do you feel subject-specific language is important to the science subject you teach your Year 10 classes? T1



Percentages may not add up to 100 due to rounding.

Figure 14: Overall, how confident do you feel you are in teaching and using subject-specific language in your Year 10 classes? T1



Percentages may not add up to 100 due to rounding.

Table 22: Current approach to subject-specific language teaching in Year 10 science classrooms, T1*

	T1		
	Intervention n (%)	Control n (%)	Total n (%)
Frequently referred to in everyday lesson vocabulary	49 (82)	36 (84)	85 (83)
Taught as a separate element in a lesson and then referred back to	5 (8)	3 (7)	8 (8)
Present in teacher assessments and formative tests only	4 (7)	1 (2)	5 (5)
Not included in your teaching at all	2 (3)	2 (5)	4 (4)
Other	0 (0)	1 (2)	1 (1)

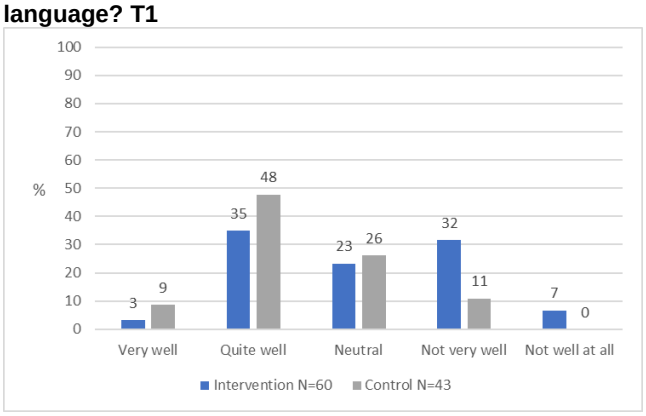
* *N = 103 (60 intervention, 43 control).*

Teachers were also asked a number of additional questions at T1 about the EAL pupils and the FLE pupils in their Year 10 classes and subject-specific language. Overall, nearly half (48% or 49 of the 103 teachers who responded to this question) felt that their EAL pupils understood the importance of subject-specific language 'very' or 'quite' well. 27% (or 28 teachers out of 103) reported that their EAL pupils understood the importance of subject-specific language either 'not very' or 'not at all' well. In contrast, 77% of respondents (or 79 teachers out of 103) reported that their FLE pupils understood the importance of subject-specific language 'very' or 'quite' well. Only 5% (5 teachers out of 103) reported that they understood the importance of subject-specific language either 'not very' or 'not at all' well.

Nearly a third of teachers (31% or 32 out of 103 teachers) reported that EAL pupils in their science classes were 'very' or 'quite' confident using subject-specific language whereas 39% (or 40 teachers out of 103) reported that they were 'not very confident' or 'not very confident at all'. 68% (or 70 teachers out of 103) reported that the FLE pupils in their science classes were 'quite' or 'very' confident using subject-specific language and 8% of respondents (or 8 teachers out of 103) reported that their FLE pupils were 'not very confident' or 'not very confident at all' using subject-specific language.

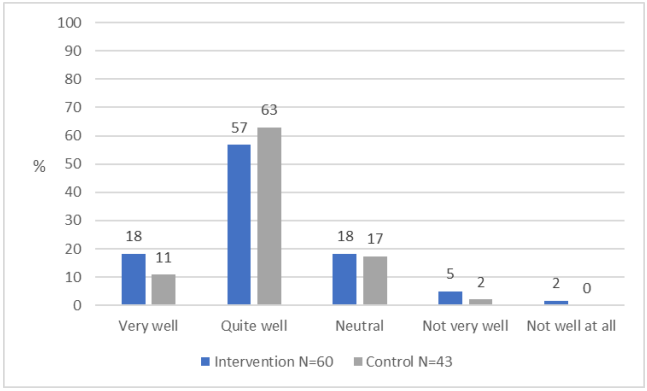
Overall, responses indicated perceived higher levels of confidence and understanding of the importance of subject-specific language and levels of engagement for the FLE pupils than for the EAL pupils. However, when the responses to these baseline survey questions are analysed by subsequent allocation (Figures 15 to 18) they suggest that teachers in the control condition were slightly more positive about their EAL pupils' confidence and understanding of subject-specific language compared to the intervention group. Although numbers are small, this may have had implications for the overall findings of the impact evaluation in terms of the potential for change as a result of the intervention. The responses for the non-EAL pupils, however, matched more closely across the two groups.

Figure 15: How well do you feel your EAL pupils understand the importance of subject-specific language? T1



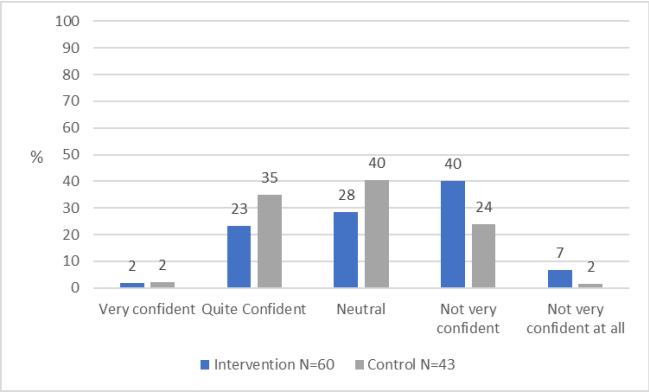
Percentages may not add up to 100 due to rounding.

Figure 16: How well do you feel your non-EAL pupils understand the importance of subject-specific language? T1



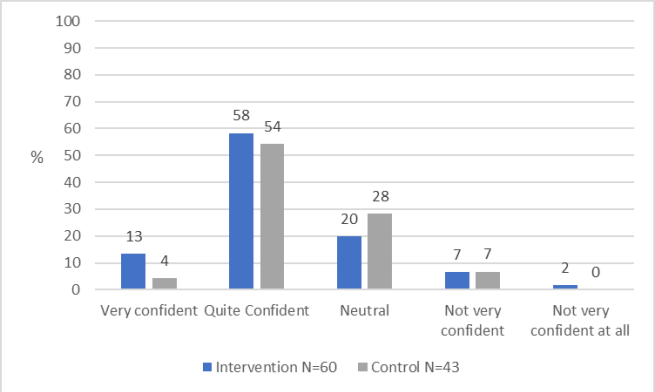
Percentages may not add up to 100 due to rounding.

Figure 17: How confident do you feel your EAL pupils are using subject-specific language? T1



Percentages may not add up to 100 due to rounding.

Figure 18: How confident do you feel your non-EAL pupils are using subject-specific language? T1



Percentages may not add up to 100 due to rounding.

Experience of training

Just under two-thirds of all science teachers in the intervention group attended all three training workshops.

- Overall, the majority of science teachers responding to the T2 survey indicated that they felt that the training provided by the workshops was effective to some degree.
- The majority of survey respondents reported completing the pre- and intersessional tasks to at least 'some' extent.
- Teachers' lack of time was a recurrent theme impacting negatively on teachers' attendance at training and completion of tasks.

In the impact evaluation compliance was measured on a continuous scale relating to attendance at training 0 = no training attended, 1 = attended one training session, 2 = attended two training sessions, and 3 = attended three training sessions. As indicated above, it became clear early on in the recruitment phase that:

- it would be difficult to recruit enough schools who could draw sufficient numbers of EAL pupils from a single science and/or history class; and
- if schools needed to involve more than one class from each subject (science/history) to meet the required EAL pupil numbers, they would potentially face a problem in releasing all the participating teachers simultaneously for the training days.

As a result, it was decided to allow a cascading process whereby one teacher from each subject would attend the training and subsequently cascade the learning to their non-attending colleagues (who were also participating in the evaluation). In addition to the training there were also a number of pre- and intersessional tasks requested of participating teachers.

This section presents findings from the monitoring data, teacher surveys, and the interview data relating to teachers experience of training, including factors impacting on attendance, whether or not cascading occurred in the event of non-attendance, and completion of programme intersessional tasks.

Attendance at training

The training was delivered in the form of three full-day training workshops delivered by each Delivery Centre to the schools within their centre. The training days were spread across three timepoints during Spring Term and Summer Term 2018—late January/early February, mid/late March and late June/early July in order to give time for teachers to embed the learning from the training and complete intersessional tasks over an extended period of time. Training was delivered by the Delivery Centres alongside a team of trainers from Challenge Partners, Hounslow Language Service, and Lampton School. If feasible, participating teachers (i.e. science and history teachers) were to be trained alongside teachers from additional subjects (not included in the trial).

All the teachers assigned to the treatment arm of the trial (N = 111) were signed up for the training—57 science teachers and 54 history teachers. Table 23 gives the attendance of the science intervention teachers from the monitoring data.

Table 23: Attendance at training*

	Monitoring data n (%)
No attendance	5 (9)
One workshop	6 (11)
Two workshops	12 (21)
Three workshops	34 (60)

* *N = 57 science teachers.*
Percentages may not add up to 100 due to rounding.

From the monitoring data, of those six teachers out of 57 who only attended one workshop, all of them only attended the first workshop. Three of these teachers were from three schools who withdrew from the intervention altogether after the first workshop (and 2 out of 57 of the science teachers who failed to attend any of the workshops were also from one of these schools).

Whilst it was hoped that the training would encompass teachers from other subject specialisms within secondary schools, attendance data indicates that, aside from 10 attendees from 4 non-participating schools (one of which was a primary school), all the workshop trainees were science and history teachers (49 of the 54 history teachers in the intervention group attended at least one workshop) or other members of staff from the participating schools. Reasons given for non-attendance at workshops in the T2 survey varied but included other school responsibilities e.g. the Office for Standards in Education, Children's Services and Skills (Ofsted), school trips, etc. (four respondents), school not being able to release multiple science teachers at the same time (two respondents), and distance/time involved (two respondents).

The workshops took place over three full days across the Spring and Summer Terms 2018. A summary of the key aims of each workshop is provided in Table 4 above in the section on 'Intervention'. The workshops were co-delivered by teachers from the Delivery Centre and trainers from Hounslow Language Service. Delivery was structured so that the teachers from the Delivery Centre delivered an increasing proportion of the material over the three workshops, which the Delivery Centre respondents felt worked well.

> *…with the 1st one, it was like diving in at the deep end…but* [the Hounslow Language Service trainer] *did most of it. Then we took it forward in the next workshop, and took on a lot more in the final one, and that definitely worked. That was a really good model.* (Delivery Centre facilitator interview)

Although the workshops were delivered within Delivery Centres, and therefore may have differed slightly according to the aspects delivered by staff from the Delivery Centres, no substantial differences in delivery were detected by the evaluation team and all workshops across Delivery Centres were structured in the same way with the same core resources provided to participants.[31]

At the end of Workshop 3, participants were asked to complete evaluation forms by Challenge Partners with questions relating to the training. A total of 35 responses were collected. Of these, 13 were history teachers and a further 7 respondents did not come from schools participating in the evaluation or were not a class teacher from one of the identified evaluation classes (e.g. they were an EAL coordinator, or Higher Level Teaching Assistant class support) so are not included here. Table 24 gives the evaluation science teachers ratings of the value of the three workshops separately, and of the training and resources provided overall.

[31] See 'Limitations' section for further discussion of (the lack of differences between) Delivery Centres in this evaluation.

Table 24: End of training evaluation forms

	Very/mostly valuable n	Somewhat/quite/fairly/moderately valuable n	Slightly/minimally valuable n	Not at all valuable n
Workshop 1*	4	9	1	0
Workshop 2*	5	9	0	1
Workshop 3**	11	4	0	0
Quality and usefulness of training and resources**	6	9	0	0

* *N = 14 (one respondent did not answer questions relating to Workshops 1 and 2).*
** *N = 15.*

Although the numbers are small it is worth reporting here as the value of the workshops appears to increase over workshops. For those who attended Workshop 3 it was perceived as very/mostly valuable by nearly three-quarters of respondents (11 teachers out of 15), with the proportions rating Workshops 2 and 1 being much lower and, in fact, lowest for Workshop 1 (5 teachers out of 14 and 4 teachers out of 14, respectively). In terms of the quality and usefulness of the training and resources overall, six teachers out of 15 regarded them as 'very/mostly' valuable and the remaining nine out of 15 rated them as 'somewhat/quite/fairly/moderately' valuable. However, it should be borne in mind that these forms were only completed by participants who attended the third (and final) workshop and therefore may show an imbalance in attitudes towards the training i.e. teachers who did not find the previous workshops useful could be less likely to attend the last workshop and therefore may be underrepresented in the third workshop survey.

Although each workshop covered both a grammatical, English language teaching element and time to introduce strategies for use with pupils (see Table 4 for the content of each workshop), the training was perceived to be delivered in such a way that it 'front-loaded' the linguistic learning elements of the programme in the first workshop, with the more practical, strategy-based elements following in the second and third workshops. The reasoning for this, as understood by a Delivery Centre representative, was *'so that, when you did the strategies later on, the foundations of language knowledge were there'.* However, this did appear to have an impact on the training:

> *Workshop 1…by the end of the afternoon the shutters had gone down on a lot of people…and then, luckily, they had a couple of months and Workshops 2 and 3 were much more user-friendly…* (Delivery Centre facilitator interview)

This was reflected in the qualitative survey responses and the teacher interviews:

> *I found the early sessions too intellectual—too focused on the grammar aspect for example. The later sessions that began to discuss implementation of the strategies were by far more useful.* (Teacher survey, T2)

Other respondents also indicated that the quantity of information provided in the training sessions was extensive, suggesting cognitive overload:

> *…so much goes in (to the training), some of the ideas, you just lose.* (Teacher interview)

It is also worth noting that much of the withdrawal and non-attendance at workshops tended to happen after Workshop 1, at which point three schools also withdrew from the programme.

Other factors, which affected the experience of the workshops included the other attendees, particularly from different subject specialisms and the difficulties of their low/variable attendance, with subsequent effects:

> *It was quite difficult because we had people from all specialities. We had scientists, we had mathematicians, we had historians, and it's very hard to kind of pinpoint an EAL need in all these subjects.* (Teacher interview)

> *[It was] very difficult with people not attending all sessions and then needing some time to go through other sessions.* (Teacher survey, T3)

As indicated above, the main reasons provided for non-attendance at training in the survey related to organisational factors, and, to a certain extent, lack of school support for programme attendance. However, in the interviews there was also an indication that pragmatic considerations were influential for the individual teachers themselves:

> *In the end it ended up clashing and the amount of work involved we [interviewee and other participating teachers at the school] couldn't see it being worth all of our time so we didn't attend the third session.* (Teacher interview)

It should be noted, that whilst all participants were aware that participation in the programme involved three full-days training, there was not the same amount of clarity regarding the extent of the intersessional tasks (see below).

Attendance and experience of training tended to also be determined by the overall perceived usefulness of the CPD provided by the programme and this is returned to in the section on 'Teacher learning' below. Overall, in the T3 survey, the majority of science teachers indicated that they felt that the training provided by the workshops was effective to some degree (32 teachers out of 38 or 84%; Table 25). However, the small number of returns at T3 may indicate that only those teachers who were most engaged with the programme completed the survey at that timepoint.

Table 25: How effective do you feel the training provided in the workshops was? T3*

	n (%)
Very effective	10 (26)
Somewhat effective	22 (58)
Neutral	2 (5)
Somewhat ineffective	1 (3)
Very ineffective	3 (8)

* *N = 38 respondents.*
Percentages may not add up to 100 due to rounding.

Cascading

As indicated above, cascading was permitted and three schools were given official permission by their Delivery Centre to do this, but this model was informally adopted in at least one other school (the idea being that either other teachers attending from that school could cascade the learning or another teacher could attend in the absent teacher's place and then cascade the training). Only five teachers (out of the 11 reporting not attending all three workshops) indicated that some cascading did occur within their school. Only two of the five reported that this cascading had been previously agreed with their Delivery Centre.

Of these five respondents, three reported they felt that all the learning had been cascaded to them whilst one indicated that they felt only some of the learning had been cascaded. The remaining teacher did not answer this question. In addition, two respondents reported that they felt the cascading process worked very well indicating that: '*I felt able to implement the learning fully, as if I had undertaken the training first hand*', two reported that it worked quite well indicating: '*I was able to implement the learning to an extent but felt at a slight disadvantage*' whereas the remaining teacher reported that the cascading process in their case did not work very well i.e. '*I struggled to implement any cascaded learning*'.[32] However, these responses should be interpreted with caution given that these were teacher self-reports (and having not attended the training they had no comparator) and there were no formal mechanisms in place for cascading.

In four schools, where the model of cascading the learning was known to have been adopted, one teacher in each school confirmed not having received the cascaded learning and not to have implemented the programme as a result. This was often because other teachers were also unable to be released from classes within school time, and any cascading that did occur was only from other teachers taking part in the evaluation who had attended the training.

[32] These were pre-specified responses in the survey question.

Inter-sessional activities

Participants were expected to complete pre-sessional and inter-sessional tasks as part of the programme. As indicated in Tables 26 and 27, not only were the tasks extensive but completion rates were low, or at least variable, although the figures are inconclusive as they come from two incomplete sources.

The monitoring data relating to the intersessional tasks indicates a completion rate of a little over 50% (or 28 science teachers out of 57) for the pre-Workshop 1 tasks (Table 26). However, there are limitations with the data with incomplete recorded completion rates. This appears to have been due to a lack of collection of such data via the Delivery Centres with Challenge Partners noting that of the 11 Delivery Centres there was no data from 4 Delivery Centres (36%, over one-third) and data for 50% of teachers was missing from a fifth Delivery Centre. Partial data also appears to be an issue for the remaining six Delivery Centres, including no information on completion of post-Workshop 3 tasks. Therefore, whilst there appears to be a decline in the number of reported completed intersessional tasks between pre-Workshop 1 and post-Workshop 1, which may reflect declining attendance and/or non-completion, it is just as likely to be a result of non-recording of completion. As Delivery Centre respondents reflected, mechanisms for monitoring and supporting, or encouraging the completion of these tasks were perhaps not sufficiently robust. Teachers were expected to upload their intersessional tasks prior to workshop attendance and Delivery Centres felt that they spent time chasing for these to be completed prior to the subsequent workshop, with little result:

> *…when they came back in the 2nd and 3rd workshops, I had to check the tasks they had done [and] half of them would say 'oh I forgot to bring it'…even though they're supposed to upload examples of everything they have done…getting hold of some evidence was very difficult.* (Delivery Centre facilitator interview)

This Delivery Centre was nominally well resourced, with dedicated admin support for the teacher trainers, but there was a perceived weakness, in that the monitoring mechanisms were not formally included in any of the roles and evidently failed to be properly picked up by anyone.

Table 26: Reported completion of intersessional tasks by participating science teachers (monitoring data)*

	Teacher task	n/N (missing)	% recorded completion
Pre-Workshop 1 tasks	Review pre-programme handbook	0/57 (57)	0
	Interviews with two Year 10 EAL pupils you teach (to share in Workshop 1)	28/57 (29)	51
	Data sheet for two EAL pupils you teach (to share in Workshop 1)	28/57 (29)	49
	Two Year 10 EAL student confidence surveys	29/57 (28)	51
	Teacher confidence survey	29/57 (28)	51
	Two subject-based student assessments: Calais (history) or Asteroids (science). Bring two copies of your student's answers and your marking to Workshop 1	29/57 (28)	51
	An example of each pupils' extended writing from a recent classroom activity or assessment. Bring two copies.	26/57 (31)	46
Post-Workshop 1 tasks	Post-Workshop 1 evaluation form	11/57 (46)	19
	Devise two activities from below (Bring soft and two hard copies):		
	Activity 1: Nominal Groups	16/57 (41)	28
	Activity 2: Building the Field	17/57 (40)	30
	Bring two copies of focus pupils' work related to activities devised above		
	Interview with focus pupils	18/57 (39)	32
	Bring two copies of lesson plan for a lesson to be taught (for use in Workshop 2)	0/57 (57)	0

Post-workshop 2 tasks	Bring a textbook for your subject (for use in Workshop 2)	0/57 (57)	0
	Self-research verb tenses in English. How do these help pupils develop fluency?	0/57 (57)	0
	Observation of/with a colleague	8/57 (49)	14
	Post-Workshop 2 evaluation form	10/57 (47)	18
	Devise two activities from below (and bring hard copies):		
	Activity 1: 1. Figurative Language; 2. Verbs; 3. Modals; 4. Nominalisation; or 5. Cohesion	14/57 (43)	25
	Activity 2: 1. Figurative Language; 2. Verbs; 3. Modals; 4. Nominalisation; or 5. Cohesion	11/57 (46)	19
	Bring two copies of focus pupils' work related to activities devised above	0/57 (57)	0
	Example of feedback: Bring two samples of feedback you have given focusing on language	0/57 (57)	0
	Bring a textbook for your subject (for use in Workshop 3)	0/57 (57)	0
	Bring an exam paper from your subject and mark scheme (for use in Workshop 3)	0/57 (57)	0
Post-Workshop 3 tasks	Post-Workshop 3 evaluation form	0/57 (57)	44
	Two student confidence surveys	0/57 (57)	0
	Teacher confidence survey	0/57 (57)	0
	Send a piece of extended writing from your two focus pupils to the central team (three–four weeks after the end of the course). Work should be marked and improvement in language highlighted	0/57 (57)	0
	Post-programme evaluation form	0/57 (57)	25

Survey responses are a little higher, albeit self-reported and from a, perhaps, more engaged cohort. However, the trend is also of declining completion of intersessional tasks from 91% (or 38 teachers out 41) of completing either all, or 'to some extent' the pre-programme tasks to 67% (or 28 out of 37 teachers) reporting completing post-Workshop 3 tasks to at least 'some extent' (Table 27).

Table 27: Reported completion of intersessional tasks by participating science teachers (survey data)[*]

	Reported completion	n (%)
Pre-Workshop 1 tasks	Yes	31 (74)
	No	3 (7)
	To some extent	7(17)
Post-Workshop 1 tasks	Yes	28 (67)
	No	4 (10)
	To some extent	9 (22)
Post-Workshop 2 tasks	Yes	21 (50)
	No	6 (14)
	To some extent	12 (29)
Post-Workshop 3 tasks	Yes	17 (41)
	No	9 (21)
	To some extent	11 (26)

[*] *Pre- and post-Workshop 1 tasks N = 41; post-Workshop 2 tasks N = 39; post-Workshop 3 tasks N = 37. Percentages may not add up to 100 due to rounding.*

Intervention teachers were asked in the Challenge Partners end of Workshop 3 evaluation forms how valuable they found these activities and the responses are provided in Table 28. It can be seen that the majority who completed these forms found the pre-workshop and intersessional activities to be of some value although they were predominantly rated as 'somewhat/quite/fairly/moderately valuable' (10 teachers out of the 15 rating the pre-Workshop 1 activities as this

and 12 teachers out of the 15 rating the intersessional in-school activities as 'somewhat/quite/fairly/moderately valuable').

Table 28: Ratings of the pre-workshop and inter-sessional activities, monitoring data*

	Very/mostly valuable n	Somewhat/quite /fairly/moderately valuable n	Slightly/minimally valuable n	Not at all valuable n
Pre-Workshop 1 activities	2	10	3	0
Intersession in-school activities	3	12	0	0

* *N = 15.*

In the T2 evaluation survey and the teacher interviews, teachers were asked about completion of these intersessional activities. Reasons for non-completion tended to focus on the amount of time required to complete them within the context of a busy teaching schedule, and a perceived lack of support for doing so:

> *It was too onerous as in like there was no incentive to be truthful from a school point of view or from a department point of view.* (Teacher interview)

> *So there was nothing in-between saying, 'By the way, this is due for four weeks' time'.* (Teacher interview)

A further factor was the feeling that, for those who completed the assigned tasks, they were not reviewed in the workshops:

> *So I thought to myself, we've done all this hard work, we done all these questionnaires, we've done all the preliminary tasks for the workshop, but it's not been acknowledged.* (Teacher interview)

Additional support

Finally, in the T2 survey, teachers were asked if they felt the need for any additional support at any time to help with their training and programme implementation. Nine teachers out of the 41 who responded to this question (or 22%) indicated that they did. Of these, one teacher stated that their needs were met and the other eight indicated that they were met 'to some extent'. Three teachers received support from the EAL coordinator in their school, two teachers received support from Challenge Partners, two teachers from their Delivery Centre, and two from another 'EAL in the Mainstream Classroom' teacher in their school. Further kinds or levels of support teachers said they would have liked to have in place included subject specialists who had already used the programme in their (science) classes and the chance to visit schools or classrooms where the EAL programme was effectively being used in different subjects.

Teacher learning

- The training appears to have impacted on teachers' support for their EAL pupils in their classrooms. At T2, teachers in the intervention group overall felt better *able* to support EAL pupils in their classroom than their control group peers and were more likely to report that they felt confident in doing so. Teachers valued the chance to better understand the needs of their EAL pupils. They also valued learning strategies that they could see were directly transferable to their classroom context. In contrast, some teachers struggled with, and failed to see the relevance of, some of the more grammatical aspects of the programme.
- Approximately two-thirds of teachers in the intervention group also felt confident in their ability to deliver the 'EAL in the Mainstream Classroom' and that the programme had changed their pedagogic practice at T2.

The aim of the programme was to enable pupils 'to develop answers, access key texts and fully understand the questions and tasks that are set' (Challenge Partners, 2017b). Consequently, there were three key areas of learning focused on in the workshops, intersessional activities, and included in the logic model: 1) understanding the complexities of academic language within teachers' subject specialism; 2) understanding the needs and diversity of EAL learners; and 3) learning and developing strategies to use in the classroom to facilitate higher levels of academic fluency, in particular

written fluency, for EAL pupils and for all pupils. These three aspects are addressed in turn before turning to overall learning from the programme.

Subject-specific and academic language

The need for subject-specific and academic language training in the programme was seen as necessary to facilitate pupils' 'capacity to develop answers, access key texts and fully understand the questions and tasks that are set' (Challenge Partners, 2017b). As indicated above, the language aspects of the programme did cause some difficulties for the participating teachers. For many it was not just about learning how to teach pupils subject-specific and academic language, including grammatical concepts, but also having to learn it themselves:

> *I struggled with the English language vocabulary, very different process from my experience.* (Teacher survey, T3)

Others struggled to see the relevance, often in terms of their perceptions of their subject specialism:

> *I feel a focus on language is not relevant to science.* (Teacher survey, T2)

Some teachers, however, welcomed the English language learning aspects of the programme:

> *We both said that actually it's a real eye-opener for us and our language skills and questioning our language skills. So, that was interesting.* (Teacher interview)

Those teachers who thought the programme was useful more generally often expressed an appreciation of the focus on literacy and vocabulary:

> *The resources and ideas were very thought provoking and made me realise the importance of literacy in science. (Teacher survey, T3)*

> *It helped with complex vocabulary in Biology.* (Teacher survey, T3)

In the T2 and the T3 survey, teachers were asked their level of agreement with the statement 'EAL in the Mainstream Classroom has improved my understanding of the linguistic demands of my subject'. Overall, 38% (or 15 teachers out of 40) at T2 strongly agreed that it had. By T3 this had decreased to 29% (or 10 teachers out of 35) strongly agreeing. However, overall the proportions agreeing 'strongly' or 'somewhat' remained relatively stable; 73% or 29 teachers out of 40 at T2 and 72% or 25 teachers out of 35 at T3 (Table 29).

Table 29: 'EAL in the Mainstream Classroom has improved my understanding of the linguistic demands of my subject'. Intervention teachers, T2 and T3

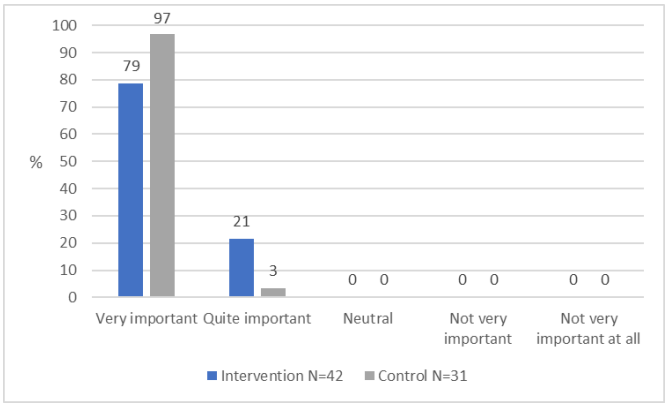
	T2* n (%)	T3** n (%)
Strongly agree	15 (38)	10 (29)
Somewhat agree	14 (35)	15 (43)
Neutral	5 (13)	3 (9)
Somewhat disagree	3 (8)	2 (6)
Strongly disagree	3 (8)	5 (14)

** N = 40 respondents.*
*** N = 35 respondents.*
Percentages may not add up to 100 due to rounding.

The T1 questions relating to teachers' rating of the importance of, and teachers' confidence in, subject-specific language were repeated at T2. Perhaps surprisingly, the proportion of intervention teachers rating subject-specific language as

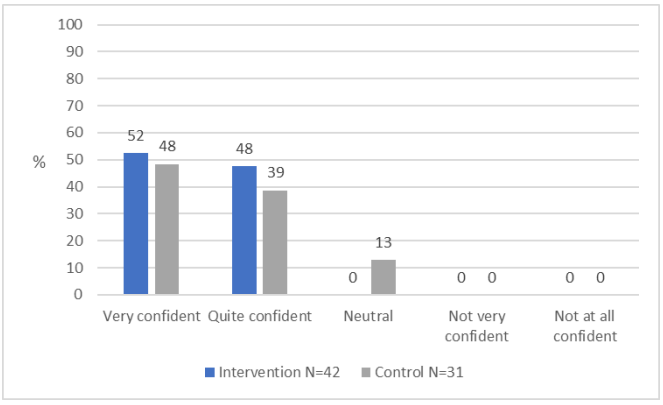
'very important' was lower at T2 than that of the control group teachers at T2 (97% or 30 of whom rated this as 'very important'; Figure 19). However, teachers in the intervention group were slightly more likely to report feeling 'very confident' at T2 (52% or 22 teachers out of 42) compared to control teachers (48% or 15 teachers out of 31) (Figure 20). However, given the small numbers these findings should be considered tentative only.

Figure 19: To what extent do you feel subject-specific language is important to the science subject you teach your Year 10 classes? T2



Percentages may not add up to 100 due to rounding.

Figure 20: Overall, how confident do you feel you are in teaching and using subject-specific language in your Year 10 classes? T2



Percentages may not add up to 100 due to rounding.

Understanding the diversity of EAL learners and their needs

For some teachers a strength of the programme was learning more about the experiences of EAL pupils. The intersessional task, which was frequently referred to in the teacher interviews was the focus on two EAL learners within their classrooms, which involved assessing their work at the start of the programme and interviewing them about their experiences, and then tracking change over time. At one level this appeared to provide teachers with a deeper understanding of particular individual children's EAL needs within a wider context, at another level it allowed a greater awareness of EAL pupils more generally:

> *But the best thing for me is having that deeper understanding of how people actually think, and how their language skills develop, and just realising how much support they need to develop properly, just so that they can access their exams.* (Teacher interview)

> *It's definitely opened my eyes about what EAL students are like and what a broad spectrum they can fit into.* (Teacher interview)

This was a lasting learning point with one respondent summing it up in the T3 survey:

> *I am more sensitive to the needs of my EAL pupils.* (Teacher survey, T3)

Strategies for use in the classroom

Part of the programme involved providing teachers with strategies and resources they could choose from and develop for use with their own pupils. These also formed part of the intersessional tasks. Programme strategies and resources included, among others, 'building the field' (e.g. pre-teaching key words), asking pupils to insert verbs into a written piece of text, nominalisation (making a noun from a verb or adjective e.g. 'diffusion' from 'diffuse'), visually organising work, making more use of text books to embed a deeper understanding of the language, use of command words to understand exam questions, the Dictogloss technique, and provision of a key word list. The interviews and survey responses (at T2 and T3) illustrated how it was important to teachers to be able to relate their training to their actual practice in schools. For some teachers the applicability of the strategies discussed was inherent in the training, both in terms of directly using some strategies or using them as ideas for their own classes:

> *The second session was much more practical based which was much more useful, and I came away with some more ideas. (Teacher interview)*

> *Due to the many resources provided, I was able to select tasks suited to my class and successfully deliver the content. (Teacher survey, T2)*

A few teachers mentioned that the content and strategies were similar to their already existing practice, although this did not appear to detract from their experiences of the training:

> *It is very similar to what I do already. However some nice new activities to try out.* (Teacher survey, T2)

However, for some teachers, the applicability of these strategies was not easily perceived and the resources not directly transferable to the classroom:

> *What people want if they're investing time (in CPD) is something to take away…the more things you can literally just pick up and drop into your teaching the better, cos we are all pressed for time. (*Teacher interview*)*

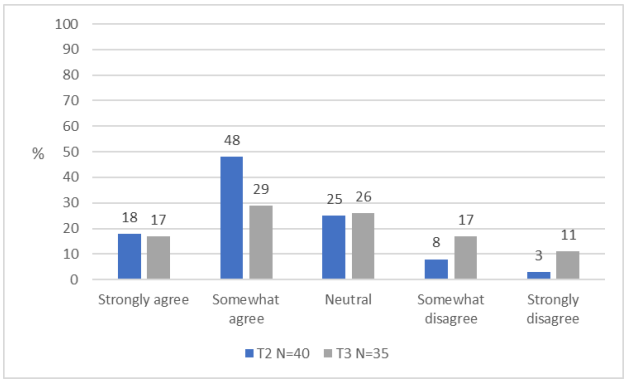
> *I felt that actual real exam questions WERE NOT USED! This led us to not clearly linking what we were learning to how to actually improve progress.* (Teacher survey, T3)

In addition, when asked about any possible changes to the training in the interviews and the T2 survey, teachers were particularly keen to be given more strategies applicable to their own classes within the sessions or to be given the time to develop their own resources based on the strategies within the sessions, rather than be built into the intersessional tasks (see above):

> *I think there needs to be something in the course where you explicitly take something and plan how you're going to execute it in the classroom before you leave.* (Teacher interview)

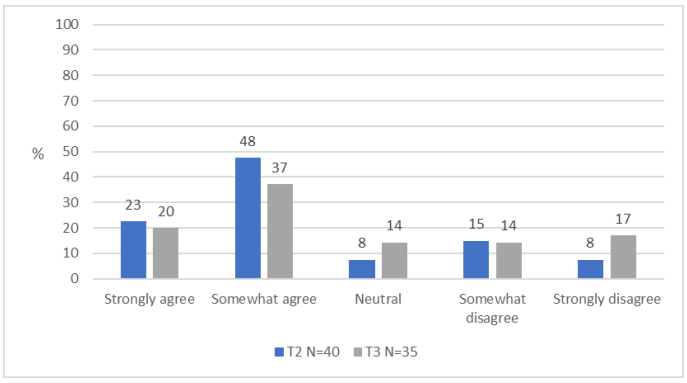
Teachers' actual use of the CPD in the classroom will be returned to in the following sections. However, teachers were asked a number of questions relating to their learning from the programme and the impact on their teaching at both T2 and T3. The majority (65% or 26 out of 40 respondents) agreed (either 'strongly' or 'somewhat') that they felt confident in their ability to deliver the 'EAL in the Mainstream Classroom' at T2 (Figure 21). This fell however, to only 46% (or 16 out of 35) respondents at T3. 70% of respondents (or 28 teachers out of 40) at T2 reported 'strongly' or 'somewhat' agreeing with the statement 'EAL in the Mainstream Classroom has changed my pedagogic practice in the classroom' and again this fell to 57% (or 20 teachers out of 35) at T3 (Figure 22). These findings are potentially a result of teachers having a year during which there was no additional input from the programme and suggestive of a dilution of implementation over time.

Figure 21: 'I feel confident in my ability to deliver the EAL in the Mainstream Classroom programme' T2 and T3



Percentages may not add up to 100 due to rounding.

Figure 22: 'EAL in the Mainstream Classroom has changed my pedagogic practice in the classroom' T2 and T3



Percentages may not add up to 100 due to rounding.

At T3, teachers were also asked to rate the usefulness of the 'EAL in the Mainstream Classroom' programme to their teaching practice. A little over half (19 out of 35 respondents or 54%) felt that it was either 'quite' or 'very' useful (Table 30).

Table 30: Teachers' rating of the usefulness of the EAL programme to their teaching practice, T3*

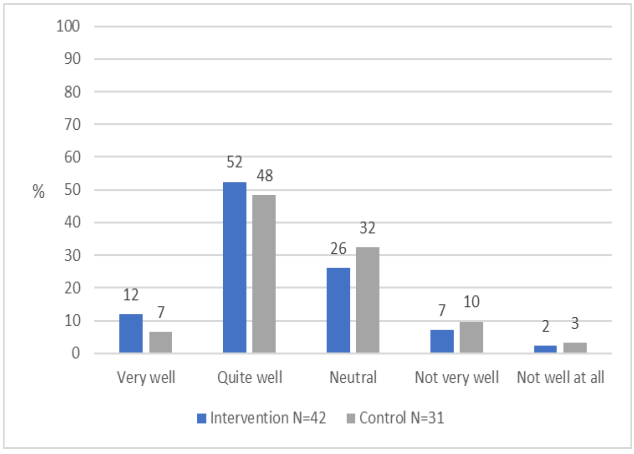
	n (%)
Very useful	7 (20)
Quite Useful	12 (34)
Useful to a small extent	9 (26)
Not very useful	1 (3)
Not at all useful	3 (9)
Didn't implement enough to decide	3 (9)

* *N = 35.*
Percentages may not add up to 100 due to rounding.

Finally, in the T2 survey, all teachers (control and intervention) were asked again about how well they felt able to support EAL children in their classroom and how confident they were in doing so in order to assess the impact of CPD on intervention teachers' teaching practice. At T2, teachers in the intervention group overall felt better able to support EAL pupils in their classroom than their peers in the control condition. For example, 12% of intervention teachers (or 5 teachers out of 42) at T2 indicated that they felt they could do so 'very well' compared to 7% (or 2 teachers out of 31) in the control group (Figure 23).

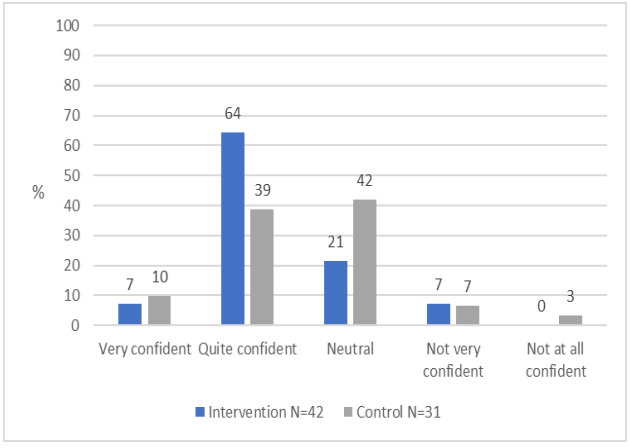
Teachers in the intervention group were more likely to indicate that they felt confident in supporting EAL pupils (either 'very' or 'somewhat') at T2 when compared to the control group (71% or 30 teachers out of 42 and 48% or 15 teachers out of 31, respectively; Figure 24).

Figure 23: Overall, how well do you feel you are able to support the EAL children in your classroom? T2



Percentages may not add up to 100 due to rounding.

Figure 24: How confident are you in supporting the EAL children in your classroom in their learning? T2



Percentages may not add up to 100 due to rounding.

Implementation fidelity

- Approximately two-thirds of survey respondents in the intervention group reported that they felt that they had been able to implement the learning from the programme either 'very' or 'somewhat' effectively at T2, falling to just over a half at T3.
- Teachers adopted those strategies they felt were most useful, could more easily be incorporated into their current lessons, and were engaging for their pupils.
- Teachers were more likely to frequently use strategies relating to knowledge checking, understanding, and (given the time of year of the observations) exam preparation. They were less likely to use strategies relating to sentence structure and grammar, which were aspects of the training teachers were most likely to struggle with and to find the least relevant to their teaching practice.
- Adaptations of the programme reported by teachers focused on the perceived needs of the class, particularly relating to their ability levels.
- Teachers reported using the programme with other year groups and sharing their learning with colleagues.

This section addresses the first part of the main IPE research question: the extent to which the programme was implemented with fidelity. As detailed above, given that 'EAL in the Mainstream Classroom' is not a manualised programme, rather CPD to aid understanding of the academic language requirements of specific subjects, needs relating to these in the classroom (specifically for EAL learners) and strategies to address those needs, fidelity is defined in this context as effective embedding of the strategies into everyday teaching with adaptation to suit individual classes (i.e. to what extent was the programme implemented). Although the focus is on participating Year 10 science classes, teachers were also asked about the use of the CPD learning with other classes and other year groups to explore the extent to which the programme had become embedded in their practice more widely. Finally, given the inherent flexibility within the implementation of learning from the CPD, this section also provides observation data to answer the second IPE question: What is the impact of 'EAL in the Mainstream Classroom' on the classroom experience?

Programme implementation

Science teachers were asked at T2 how effectively they felt they were able to deliver the 'EAL in the Mainstream Classroom' programme to their participating classes. A total of 26 science teachers (out of 40 or 65%) felt that they had been able to implement the learning either 'very' or 'somewhat' effectively during the first year of the trial (i.e. at T2). At T3 51% (or 18 out of 35 teachers) perceived that they implemented the 'EAL in the Mainstream Classroom' ('very' or 'somewhat') effectively. There was little difference in the findings when only those respondents who answered at both timepoints were included in the analysis (Table 31).[33] However, at T3, seven respondents indicated that the question was 'not applicable' and where a reason for this was given it was because they did not deliver the programme in the academic year 2018/2019.

Table 31: Intervention teachers perceived effectiveness in delivering the 'EAL in the Mainstream Classroom', T2 and T3*

	All respondents at T2 and T3		Responded at both, T2 and T3***	
	T2* n (%)	T3** n (%)	T2 n (%)	T3 n (%)
Very effectively	6 (15)	6 (17)	4 (13)	5 (17)_
Somewhat effectively	20 (50)	12 (34)	15 (50)	9 (30)
Neutral	6 (15)	_****	4 (13)	_****
Somewhat ineffectively	6 (15)	7 (18)	5 (17)	7 (18)
Very ineffectively	2 (5)	3 (8)	2 (7)	2 (7)
Not applicable	0 (0)	7 (18)	0 (0)	7 (18)

* *N = 40.*
** *N = 35.*
*** *N = 30.*
Percentages may not add up to 100 due to rounding.
**** *This option was not given at T3.*

[33] These two findings are, not, however, directly comparable given at T2 this was asked as a 5-point scale and at T3 as a 4-point scale.

Teachers were asked again, at T2, about their approach to subject-specific language teaching in their Year 10 science classrooms. A higher proportion of intervention teachers reported teaching subject-specific language as a separate element in a lesson and then referred back to whilst teaching the lesson content compared to control teachers (17% or 7 teachers out of 42 and 10% or 3 teachers out of 31, respectively; Table 32) and this was higher for intervention teachers at T2 than at T1 (8% or 5 teachers; Table 22). However, as at T1 the majority of teachers in both groups stated that subject-specific language was frequently referred to in everyday lesson vocabulary (83% of all teachers at T1 or 65 teachers out of 103 and 84% or 61 teachers out of 73 at T2). It should also be noted that the proportion of control teachers stating that they frequently referred to subject-specific language teaching in their everyday lesson vocabulary at T2 was higher when compared to those in the intervention group (28 teachers or 90% compared to 33 teachers or 79%). It is difficult to interpret these findings given that the programme aims to embed strategies within classroom teaching. Although the implication is that, for some 'EAL in the Mainstream Classroom' intervention teachers, more distinct time is given to teaching subject-specific vocabulary and grammar, yet the sample size and the differences between the two groups are small.

Table 32: Teachers' approach to subject-specific language teaching in Year 10 science classrooms at T2*

	Intervention n (%)	Control n (%)	Total n (%)
Frequently referred to in everyday lesson vocabulary	33 (79)	28 (90)	61 (84)
Taught as a separate element in a lesson and then referred back to	7 (17)	3 (10)	10 (14)
Present in teacher assessments and formative tests only	2 (5)	0 (0)	2 (3)
Not included in your teaching at all	0 (0)	0 (0)	0 (0)
Other	0 (0)	0 (0)	0 (0)

* *N = 73 (42 intervention, 31 control).*
Percentages may not add up to 100 due to rounding.

In the face-to-face interviews, teachers were asked what they had found useful and/or challenging about applying the learning to their practice. Those teachers who commented on aspects of the programme that they had found useful focused on very specific elements:

> *The structures I've used in the past are key words…but I think breaking it down to the structure of the sentence and how the words relate has been very useful.* (Teacher interview)

> *I use the vocabulary building strategies often due to the subject being rich in this area. I use grammar and the use of building descriptive sentences very little, as this skill does not help them to gain exam marks.* (Teacher survey, T3)

Overall, there emerges a picture of teachers trying some of the strategies, some of the time, based on what they felt were useful, easy to implement, and engaging for the pupils in their class:

> *There were only a few strategies that I could use constantly, while others were difficult in terms of pupil engagement and application to science lessons.* (Teacher survey, T2)

> *The ideas that came from the training sessions were fairly easy to implement'* (Teacher survey, T2). (Teacher interview)

For some teachers, implementation tended to take place either just after a workshop, or just before. However, for others the programme became more embedded in their practice:

> *I am aware of the training that I have undergone, and therefore I think about use of language etc as I plan and deliver my lessons.* (Teacher survey, T3)

Where implementation was deemed to be challenging time pressures were seen as barriers to effective implementation, both in terms of the teachers getting on top of the learning themselves, and also trying to incorporate it into their teaching practice:

> *I've had no time to actually think oh when can I use it here, when can I use it there…I think it was too much information for me to process in that time, and then go and do it.* (Teacher interview)

Strategies were great—time is the greatest factor. (Teacher survey, T3)

Consequently, whilst as mentioned above, some teachers felt that they could easily use strategies and resources with their classes, others felt that they needed more time to prepare resources and would have ideally liked some of that time to have been during the training workshops themselves.

Resources had to be made from scratch which is very time-consuming. (Teacher survey, T2)

I think it could have [had]…importantly, time scheduled in for us to actually make some resources for upcoming lessons. (Teacher survey, T2)

This was exacerbated by a change in the Key Stage 4 curriculum at the time of the evaluation, which meant that schemes of work had to be re-written, but with no necessity for an increased focus on language:

In the old spec there used to be much more emphasis on English and quality of written communication it was called. (Teacher interview)

At the end of the day most of our marks are on the scientific content not on how they've written it but I do understand we've got to get all our children regardless, EAL or not, we've got to get them up to the academic language. (Teacher interview)

Adaptations

Out of the 40 teachers who responded to the survey question on whether or not they made any adaptations to the programme at T2, 27 teachers (68%) stated that they did and 13 teachers (33%) stated that they did not. However, as indicated above, the programme was designed to be adapted depending on the teacher's assessment of their pupils' needs. Consequently, when teachers discussed adaptations (in both the open-ended survey questions and in the interviews) they predominantly focused on the ability levels of the pupils:

Five a day activities were adapted in order to make them sufficiently challenging for the target group. Nominalisation tasks were also used in the same way. (Teacher survey, T2)

If I get a bottom set next year, I'd need to use the strategies in a different way. I'd need to think more carefully, and maybe slow things down a bit more, so it might be more about getting them to link sentences with connectives. That would take them a lot longer. (Teacher interview)

These examples are where teachers saw that there was the potential for differentiation within the programme. Others, however, perceived that the programme wasn't suitable for some (usually high ability) classes:

Not applicable to my EAL learners, they are far too advanced and just fall into that category. (Teacher survey, T2)

In the interviews, the actual personalities and composition of the class was also talked about more, usually as a result of the preceding lesson observation:

Things might work well with a certain group but might not work with another. (Teacher interview)

Implementation outside of participating Year 10 science classes

Teachers were expected to embed the learning from the 'EAL in the Mainstream Classroom' programme within their practice, which would enable them to continue to use them with their Year 10 science pupils when they moved into Year 11. Consequently, intervention teachers were asked in the T3 survey if they were implementing the EAL programme with any classes/year groups other than the classes participating in the trial and, if so, the extent to which they were implementing it. The expectation was that teachers would continue to use programme strategies and learning with their subsequent Year 10 classes and with other (non-trial) classes and year groups. Of those that indicated that they were implementing this learning with other classes in the second year of the evaluation (21 teachers), a third of respondents (7 teachers out of 21 or 33%) indicated that they did so 'to a great extent' (Table 33).

Table 33: Extent of second year EAL programme implementation with non-trial classes, T3*

	n (%)
To a great extent	7 (33)
To some extent	9 (43)
To a small extent	5 (24)

* *N = 21.*

From the detail these teachers provided (N = 21), the programme was being implemented with a wide range of other year groups and classes. As one teacher stated:

> *I have found that the strategies work with lots of classes not just the particular Year 10 group I started with.* (Teacher survey, T3)

Intervention teachers were also asked (at T3) if they thought that the programme would benefit from being introduced to a different age group. Of those who suggested what age that should be (N = 17), 11 (65%) picked the beginning of secondary school/Year 7, with 4 out of the 17 suggesting earlier (i.e. primary), and 2 suggesting Year 8 (Table 34). Essentially, the view was that it would be beneficial to introduce the programme much earlier than Year 10.

Table 34: Applicability of 'EAL in the Mainstream Classroom' programme to other age groups, T3*

	n (%)
Year 7/Start of Secondary School	11 (65)
Primary School	4 (24)
Year 8	2 (128)

* *N = 17.*
Percentages may not add up to 100 due to rounding.

In addition, teachers were asked if they would continue to use the learning from the CPD in their future teaching practice to gain some insight into the extent to which the 'EAL in the Mainstream Classroom' learning had become embedded. Almost two-thirds of those who answered this question at T3 (21 out of 33 or 64%) reported that they would definitely be implementing the learning from the programme in their future teaching practice (Table 35).

Table 35: Likelihood of implementing the EAL programme learning in future teaching practice, T3*

	n (%)
Definitely will continue to apply the learning	21 (64)
Unsure	5 (15)
Unlikely to continue applying the learning	7 (21)

* *N = 33.*

Alongside this, several teachers indicated that they had fed back learning from the training workshops either through work with the EAL coordinator or through in-service training day (INSET) sessions. Two teachers mentioned the in-school CPD they provided as being part of a 'whole-school approach'.

'EAL in the Mainstream Classroom' in the classroom

The live observations conducted by researchers in the classroom were chosen to represent a variety of teachers from different geographical locations and Delivery Centres The observations took place over the whole lesson period and aimed to observe classroom behaviours, the presence of any of the learned 'EAL in the Mainstream Classroom' strategies and the activities undertaken in order to understand 'the classroom experience'. Six observations are reported

on here.[34] When interpreting these it should be noted that they were a snapshot of one lesson in a particular period of time (Summer Term 2018) and different teacher and student behaviours can, understandably, be expected depending on the type and content of the lesson being taught. In addition, although the observations were non-participatory there would have been an awareness by teachers and pupils of the researcher's presence in the classroom. It should also be noted that as the observations took place during the Summer Term (in order to allow some embedding of the programme to have taken place) some classes were preparing for, or in one case undertaking, assessments.

Of the six lessons observed, three were chemistry classes, two were biology classes, and one was a physics class. However, the exact qualification (i.e. whether the three separate science GCSEs or as part of combined science GCSEs) was not generally ascertained.

Two observer checklists were used in each observation designed to capture: 1) general classroom behaviours as the strategies are meant to be embedded into normal classroom practice rather than taught separately; and 2) frequency and type of 'EAL in the Mainstream Classroom' strategies used during the lesson.

For the first checklist, observers were looking for certain classroom behaviours, by both teachers and pupils. The items were chosen to assess whether or not this was usual practice for a lesson (item 'pupils are familiar with tasks') and to assess teaching behaviour (e.g. use of extension activities and differentiation), pupil engagement, and student confidence. The findings are provided in Table 36 below. It is notable that there was little provision of extension activities or differentiation. Although it is difficult to determine if this was a result of the programme being a 'whole-class' intervention or, as is more likely, due to the pupils being set by prior attainment levels. The item 'tasks assigned to pupils are context embedded' aimed to assess whether the academic language-related activities relating to the programme were embedded within the context of the science topic being covered. In addition, the 'use of visual aids' was included as an item as this was included in the training to aid students in ordering and organising written answers, although it is also a common feature within science classes more generally.

Table 36: Observed classroom behaviours

	Observed* N	
	Often	Sometimes
Teacher provides extension activities / differentiates	1	1
Pupils contribute to the lesson	2	3
Pupils are engaged with the lesson topic and on-task	2	3
Pupils are confident in discussing the topic	2	2
Pupils are familiar with tasks	1	3
Tasks assigned to pupils are context embedded	3	2
	Yes	**No**
Visual aids used e.g. diagrams, charts, video, photos	4	1

** N = 6 lessons. If a behaviour was not observed or not applicable to the lesson taught it was not noted down on the checklist.*

The second observation checklist was used to identify which areas of learning were focused on within the lessons. The activities the checklist focused on were all associated with strategies and techniques emphasised within the 'EAL in the Mainstream Classroom' CPD. Table 37 provides a summary of the observed activities. As can be seen checking knowledge and understanding was the more prevalent activity, used in most lessons and with most frequency. As could be expected, given the year group (Year 10) and time of year (Summer Term) a great deal of emphasis was also placed on exam preparation. In contrast, activities associated with sentence structure and grammar were less prominent.

[34] The seventh observation was from a school, which later did not sign the amended data sharing agreement and subsequently had to be excluded from the analysis.

Table 37: 'EAL in the Mainstream Classroom' strategies observed

Activity	Observed*	
	Number of classes	Frequency
Knowledge/understanding checking (e.g. building the field)	5	18
Exam preparation (use of command words)	4	12
Academic language/extending vocabulary (e.g. key words, nominalisation)	4	9
Sentence structure (e.g. extended sentences, use of connectives)	4	6
Grammar skills (e.g. uses of verbs, nouns)	1	1

* *N = 6 lessons.*

Finally, the researchers made detailed notes on the lessons and the exact strategies used. The lessons were then graded as either at 'emerging', 'establishing', or 'embedded' levels of implementation of the programme in the classroom. Vignettes of what the programme looked like in the classroom are provided below (Examples 1–6), based on the observation notes and supplemented, where appropriate by some data from the teacher interviews referring to the observed lesson. It should be noted that the number of programme-related strategies used in a particular lesson did not necessarily equate to whether or not programme implementation was regarded as emerging, establishing, or embedded within the lesson. For example, in Example 6 (classed as embedded), the teacher used a wide variety of strategies in one lesson with a top set biology class. In contrast, in Example 5 (also classed as embedded), the teacher used fewer strategies from the programme but did so in more depth with the focus being on checking understanding, extending vocabulary, and sentence structure with a class he described as '*challenging*': '*They struggle to write, they struggle to form their ideas, although they're bright*'. As above, it should also be noted that these examples are a snapshot of practice in a particular lesson with a particular class at a particular point in time and some lessons/topics may have lent themselves more readily to adoption of programme strategies than others.

Emerging

Example 1: Chemistry exam preparation

This was a 55-minute combined science revision lesson. Most of the lesson was spent going through the pupils' mock exam results and their written mock examination papers. Teacher and pupils went through each exam question alongside the mark scheme. In doing so, the teacher was knowledge checking, checking understanding, and ensuring the use and importance of academic language. As they went through the exam papers, pupils were asked to highlight words they were unfamiliar with. The teacher then went through a couple of the highlighted words at the end of the lesson e.g. 'penetrating' and 'ionising', checking the pupils knew the correct spelling and meaning. This activity was to continue in the next lesson (which was scheduled for after lunch the same day, this lesson having taken place in the period just prior to the lunch break). The pupils were engaged and on-task throughout the 55-minute lesson and were confident in discussions.

Example 2: Chemistry revision

This was a high-ability class taking a one-hour chemistry revision lesson. The lesson started with a '5 a day' activity (five questions on the board) that the pupils were asked to answer from memory. This appeared to be the usual starter for lessons. The pupils then worked through the '5 a day' questions in a revision booklet. The teacher worked through the answers to the questions, on the interactive whiteboard (IWB) and engaged with the pupils to explain these answers. The pupils marked their own work. The teacher also related the exercise to what areas the pupils needed to revise. Pupils were then asked to complete a matching activity using a table of properties. These activities all took place during the first 30 minutes of the lesson. During the second half of the lesson, the teacher provided the pupils with a second booklet to complete, which comprised of a revision sheet required for their GCSE practicals. Whilst the pupils worked through the booklet, the teacher circulated the class, often engaging with the class as a whole and sometimes with small groups and individuals. Extension activities were also provided. In the subsequent interview, the teacher explained that the '5 a day' was a lesson starter advocated by the school she had adapted to use 'command words' after taking the 'EAL in the Mainstream Classroom' programme training.

Establishing

Example 3: Biology – ecology assessment

Pupils had experienced three previous lessons revising this topic and the current lesson was an assessment of their learning. There were sentence starters in five different languages displayed on the walls. The teacher stated that they used these displays in other lessons as well. He also provided his EAL pupils with dual language dictionaries. The pupils were then settled down to proceed with the assessment under exam conditions. The teacher provided one to one explanations when needed to check pupils' understanding. After 30 minutes of the one-hour lesson the teacher selected Question 4 on the assessment and talked through it with the whole class as it was felt to be unfamiliar to the pupils. He explained the marking scheme and what was needed for each mark. After ten minutes the pupils continued with the assessments individually until the end of the lesson. In the interview afterwards, the teacher talked about how he now designs assessments: '*I have to design in, here's some tricky bits, and just sort of increase the cognitive difficulty and the amount of reading they need to do as they move through*'. He also discussed how his expectations of his pupils had become higher than they would have been prior to the 'EAL in the Mainstream Classroom' training: '*So, the last question generally would be a sort of a six marker, but I would kind of understand that most of the kids wouldn't do it. It was almost there to differentiate for the more able students…Whereas now, my expectation is they all do that'.*

Embedded

Example 4: Biology revision lesson – transport in animals and plants

The teacher used a variety of strategies to engage the pupils in their learning, starting with 'building the field' and ending with a structure for a longer (6 marks) exam question This involved:

- labelling diagrams (heart and lungs) whilst checking spellings;
- a matching activity relating to descriptions and explanations;
- looking at command words in exam questions, with reference to the marking scheme, followed by the pupils writing their answer to the question based on the matching activity;
- pupils annotating a 6-marks exam question and discussing key words to use in the answer; and
- pupils planning their answers individually using the structure provided by the teacher, including a table with two columns relating to the circulatory system. Pupils were to complete the second column.

The pupils were engaged and appeared confident in discussing the topic and the teacher engaged frequently with the class as a whole. In the post-lesson interview, the teacher was very positive about the training and how applicable the strategies, including the matching activity, were: '…*it's something you can go away and actually use, you don't have to sit down for an hour thinking 'how do I actually apply this?'*

Example 5: Physics lesson—forces

This hour-long lesson involved predominantly experimental work on how to measure the force on a current carrying wire on a magnetic field. The lesson started with a discussion of forces ('building the field') including partner talk to clarify terms (e.g. greater, correlation, directly proportional). The teacher then checked the pupils' understanding of maths terms and how to apply them in science. Pupils were asked to write a sentence using the following words: force, greater, correlated, and directly proportional. The teacher named the equipment as he demonstrated how to set up the experiment. The pupils then worked in pairs to conduct the experiment. Whilst the pupils were tidying up after the experiment, the teacher started to talk about verbs and nouns and how they link together. After the pupils were all settled the teacher indicated some nouns and verbs (which were labelled as such) on the IWB. He asked the pupils to formulate a sentence using these nouns and verbs to explain what they had observed in the experiment. In the subsequent interview the teacher indicated the most useful learning points for him from the 'EAL in the Mainstream Classroom' training were using the textbooks more and paying more attention to verbs and nouns (as just demonstrated in the lesson). One of the changes he has noticed in his own practice was rather than just focusing on key words, which he had done previously, he found: '*breaking it down to the structure of the sentence and how the words relate has been very useful*'. He expanded on this later in the interview: '*It's always the relation and science*

key words, you always think they're the most difficult things, so nucleus or positive charge. But in some ways its emitted, reflected and deflected are the crucial bits to understand what's happening to those.'

Example 6: Chemistry lesson—ionic compounds

This lesson began with a matching activity, using connectives. The lesson also involved a lot of work around organising knowledge, for example through looking at different ways of grouping (i.e. bonding of ionic compounds) and a decision tree (type of binding dealing with), use of key words (ionic, covalent, metallic, and giant) and checking understanding of these words. Pupils were encouraged to develop their own decision tree using an example provided on the IWB. There was also a focus on how this learning could be applied to exam questions, both 3-marks questions and the more extended writing 6-marks questions. In her interview the teacher indicated that Dictogloss and active listening activities had also been planned but not delivered due to time pressures. She also explained:

> *I've always focused on keywords and things for science, because they're quite key in the exams, you know, to make sure the kids are using the correct terminology and those type of things. But I hadn't really thought about how you link those things or how you think a little bit further…I've made a big point of going through looking at command words in exams and getting them to think about what does that actually mean?*

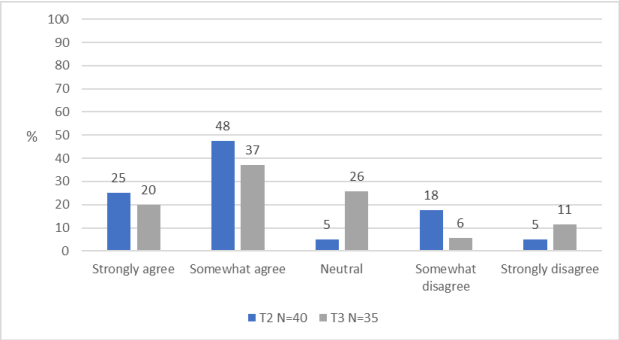
Perceived impact of the programme

- Overall, teachers in the intervention group were positive about the potential of the programme to impact on all pupils learning (both EAL and FLE pupils).
- Positive impacts were particularly seen for EAL pupils with teachers in the intervention group more likely than those allocated to the control group to report (at T2) that their EAL pupils were confident in using subject-specific language and that they were 'very well' engaged in their science lessons.
- Positive impacts on FLE pupils were, however, more mixed. Whilst teachers allocated to the intervention were more likely (at T2) to report that they FLE pupils were 'very well' engaged in their science lessons, than teachers allocated to the control group, they were less likely to report that their FLE pupils were confident using subject-specific language.

This section addresses teachers' perceived impact of the programme on the classroom, with particular regard to pupils' learning.

In the T2 and T3 surveys, teachers were asked about their confidence in the programme's ability to impact positively on pupils' learning. As seen in Figure 25, at T2 the majority of respondents agreed (either 'strongly' or 'somewhat') with this statement (73% of respondents or 29 teachers out of 40). However, this was lower at T3 with only 57% (or 20 teachers out of 35) 'strongly' or 'somewhat' agreeing and the number 'strongly' disagreeing doubling from 5% (or 2 teachers out of 40) at T2 to 11% (or 4 teachers out of 35) at T3.

Figure 25: I have confidence in the EAL in the MC programme to impact positively on EAL student outcomes, T2 and T3

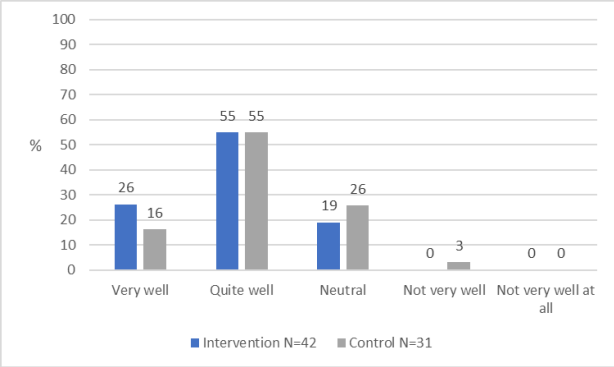


Percentages may not add up to 100 due to rounding.

All teachers (control and intervention teachers) were asked again at T2 about the levels of student engagement in their Year 10 classes, about their pupils understanding of the importance of subject-specific language and their confidence in using it. As at T1, they were asked these questions for both EAL and FLE learners separately. The responses for each question are presented in Figures 26–31 and summarised below:

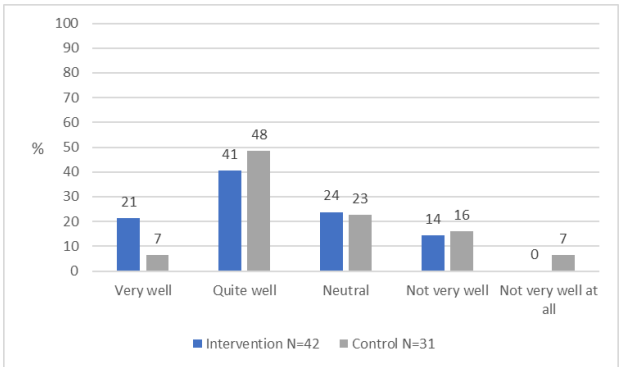
- At T2, 26% of intervention teachers (or 11 teachers out of 42) stated that their EAL pupils and their FLE pupils were ‘very well’ engaged compared to 16% of those in the control group (or 5 teachers out of 31).
- 21% of intervention group teachers (or 9 teachers out of 42) stated that their EAL pupils understood the importance of subject-specific language ‘very well’ at T2 compared to 6% (or 2 teachers out of 31) in the control group.
- At T2, 55% of intervention teachers (or 23 teachers out of 42) rated their EAL pupils as ‘very’ or ‘quite’ confident in using subject-specific language compared to 48% of those in the control group (or 15 teachers out of 31).
- At T2, proportionately more intervention teachers also reported that their FLE pupils understood the importance of subject-specific language ‘very’ or ‘quite’ well, compared to teachers in the control group (91% or 38 teachers out of 42 and 78% or 24 teachers out of 31, respectively).
- In contrast, 77% of teachers in the control group (or 24 teachers out of 31) at T2 rated their FLE pupils as ‘very’ or ‘quite’ confident in using subject-specific language compared to only 67% of the intervention teachers (or 28 teachers out of 42).

Figure 26: How well do you feel your EAL pupils are engaged during your lessons on the whole? T2



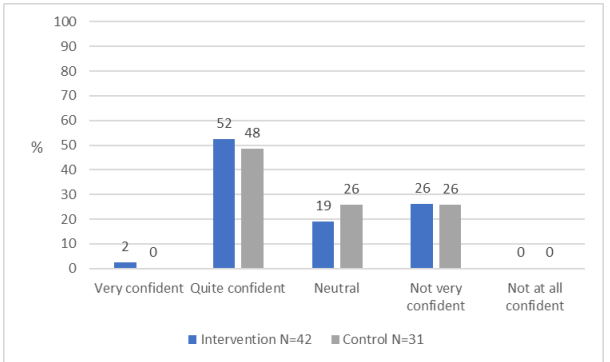
Percentages may not add up to 100 due to rounding.

Figure 27: How well do you feel your EAL pupils understand the important of subject-specific language? T2



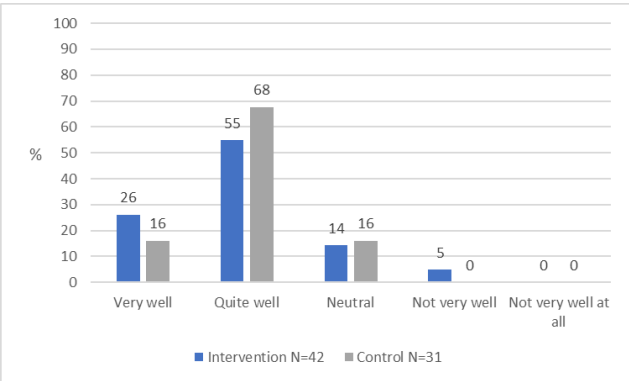
Percentages may not add up to 100 due to rounding.

Figure 28: How confident do you feel your EAL pupils are using subject-specific language? T2



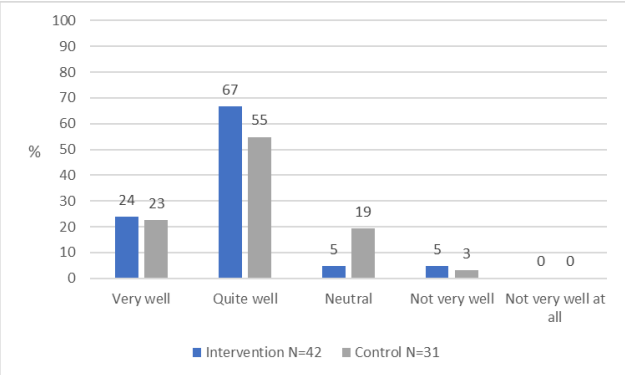
Percentages may not add up to 100 due to rounding.

Figure 29: How well do you feel your non-EAL pupils are engaged during your lessons on the whole? T2



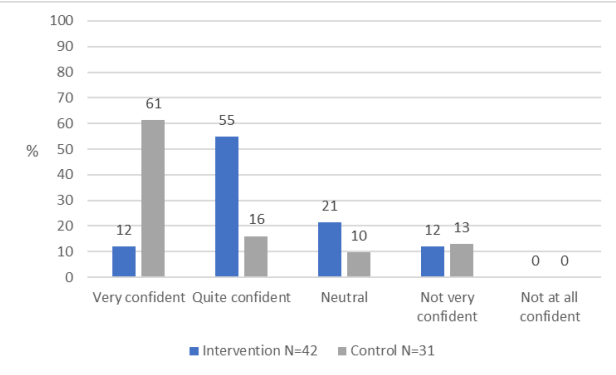
Percentages may not add up to 100 due to rounding.

Figure 30: How well do you feel your non-EAL pupils understand the important of subject-specific language? T2



Percentages may not add up to 100 due to rounding.

Figure 31: How confident do you feel your non-EAL pupils are using subject-specific language? T2



Percentages may not add up to 100 due to rounding.

In the face-to-face interviews, five teachers provided their perceptions of the impact of the programme on their pupils. Some teachers also provided additional comments in the T2 survey. Positive comments emphasised pupil engagement in particular:

> *I feel the students engaged well with the tasks, particularly as they got immediate feedback on tasks relating to correct tenses/choosing the correct word to make a sentence sound correct. This is referring to all students, not just EAL students.* (Teacher survey, T2)

> *My students are looking more confidently at textbooks…I can point to (pupil name) being able to formulate better writing…I wonder whether he would have been able to do that if I hadn't structured it with nouns and verbs.* (Teacher interview)

Whilst, as seen above, some teachers felt that the programme was not suitable for their higher ability learners, others did find it useful, even with those groups of children:

> *Some of the stuff like the verbs and the nominals and that, we did that later. It was quite interesting, they've actually struggled with that* [during the intervention] *even though they're bright kids, they found that quite difficult.* (Teacher interview)

Many teachers in the interviews and survey responses also indicated, as posited in the logic model, that the programme was not just useful for EAL learners:

> *And it's not necessarily that, it's about for students who don't have a background in very formal English, which includes pretty much most of our White British kids.* (Teacher interview)

> *Kind of all the pieces fit into place. It was like actually, if you do X, Y and Z it supports everybody*. (Teacher interview)

However, as seen in the section on 'Programme implementation' above, and further examined in the section on 'Dosage' below, some teachers reported not having time or other pressures impacting on the extent to which they were able to implement:

> *[M]ore time is required to embed strategies.* (Teacher survey, T2)

This had subsequent effects on any perceived outcomes:

> *As I mentioned I did the odd activity with the class but need to do more for it to make an impact.'* (Teacher survey, T2)

Dosage

Over two-thirds of survey respondents indicated at T3 that they were teaching the same class(es) as they had done the previous year (i.e. the same Year 10 pupils in Year 11).

However, less than a third indicated that they continued to use learning from the programme 'to a great extent' and approximately a fifth stated that they no longer used the programme at all (although it should be noted that the overall number of responses was low [n=37]).

In terms of impact and implementation, dosage is an important factor. For the purposes of this report, dosage is defined as the amount of time intervention group teachers spent teaching their Year 10 science classes. This is because the programme is a series of strategies learned through CPD for teachers to apply, as needed, in their own teaching rather than a manualised programme or a series of activities that should be undertaken in order to be effectively teaching using the programme learning. These strategies and their use are discussed more fully in the section on 'Implementation fidelity' above.

As seen in the 'Reach' section above, the majority of classes taught by intervention teachers were taught in at least a third of their science lessons by an 'EAL in the Mainstream Classroom' intervention teacher (i.e. at least all of the biology, chemistry, or physics content as part of the combined science GCSEs option, or all of a single science subject for those pupils taking the three separate science GCSEs). Given the nature of the programme we assume that this equates to

the amount of time classes benefited from a teacher's learning from the CPD, although we recognise that this is an imprecise proxy measure. Whilst at first glance this does not seem like a high proportion of time spent teaching intervention classes this should, however, be placed within the context of the amount of total science teaching for any pupil at GCSE level, which would be at the equivalent of that provided for two or three GCSEs. Therefore, this would equate to approximately at least the equivalent of two-thirds of a GCSE (i.e. a third of two science GCSEs for combined science candidates) or one GCSE for those studying for the three separate sciences. Over a quarter were taught all three science subjects by the same teacher during Year 10 (i.e. three GCSEs or, more likely, two GCSEs) (Table 17).

A further factor was whether or not the teacher continued to teach the same class in the second year of the trial and if they did so whether or not they implemented the learning from the CPD. In the T2 survey, teachers were asked whether they expected to be teaching their Year 10 classes the following academic year (i.e. 2018/2019), when the pupils were in Year 11. Less than two-thirds of all responding teachers (61% or 44 out of 72) indicated that they expected to be teaching the same classes. For the intervention group this was lower, at 59% or 24 teachers out of 41 with a further 17% unsure (7 teachers out of 41) (Table 38). At T3 intervention teachers were asked if they were, in fact teaching the same class(es) (i.e. the same Year 10 pupils in Year 11) and over two-thirds (68% or 25 teachers out of 37) indicated that they were (Table 39).

Table 38: Expectations of teaching current Year 10 science classes the following academic year (2018-2019), T2*

	Intervention n (%)	Control n (%)	Total n (%)
Yes	24 (59)	20 (65)	44 (61)
No	10 (24)	9 (29)	19 (26)
Unsure	7 (17)	2 (7)	9 (13)

* *N = 72 (intervention = 41, control = 31).*
Percentages may not add up to 100 due to rounding.

Table 39: Teaching the same classes 2018-2019, T3*

	Intervention n (%)
Yes	25 (68)
No	12 (32)

* N = 37.

Those that said they were not teaching the same classes cited reasons concerning timetable changes (n = 6), pupils changing classes (n = 3), and staffing/set changes (n = 3).

Those who were teaching the same participating classes in Year 11 were asked about the extent to which they felt they were able to implement the programme. Less than a third of teachers who responded to this question (29% or 7 out of 24 teachers) stated that they did so 'to a great extent' and just over a fifth (21% or 5 teachers out of 24) indicated 'not at all' (Table 40).

Table 40: Extent of second year programme implementation among those teaching the same classes, T3*

	n (%)
To a great extent	7 (29)
To some extent	7 (29)
To a small extent	5 (21)
Not at all	5 (21)

* *N = 24.*

Those who did state they continued to implement the programme stated that they did so because of the positive impact they perceived as a result of the programme. This second year was also a time to revisit and further embed the programme strategies, which was an ongoing process:

> *For me to fully change my teaching strategy, it would take time to build the habit up.* (Teacher survey, T3)

Those who did not implement the programme, or only implemented it to a small extent, often cited practical barriers (particularly time pressures), a whole-school approach taking precedence, or they did not perceive the programme to be sufficiently useful to continue using:

> *I could have used it more if it was higher up the list of priorities!* (Teacher survey, T3)

In addition, teachers were asked in the T3 survey if they were likely to implement the learning in their future practice. 21 teachers (or 64%) out of the 33 who responded to this question stated that they 'definitely' would implement the programme in the future, five teachers (or 15%) stated that they were unsure, and seven teachers (or 21%) stated that they would not.

Cost

The evaluation team collected cost data from Challenge Partners, supplemented with additional data from the intervention teacher interviews with the intent of establishing the average costs of implementation of the 'EAL in the Mainstream Classroom' programme in this trial. Although this evaluation was set up following the EEF earlier cost guidelines (EEF, 2016), we have presented the costs bearing in mind the updated guidance (EEF, 2019). We estimate average marginal costs per pupil per year for schools implementing the 'EAL in the Mainstream Classroom' programme. The costs over a three-year period have also been estimated.

Financial costs

The main financial costs for this trial were implementation costs (i.e. training the evaluation schools) and additional costs for schools, which primarily involved staff travel to attending training.

Although the trial took place over a two-year period the delivery of the training to the schools took place in the first year. Financial costs are divided below into delivery costs (for the training) and schools costs (attendance at training) all of which are regarded as set-up costs. Additional school costs of implementing the training within schools are also discussed.

Delivery costs

The delivery costs were originally budgeted for 12 Delivery Centres, providing three training workshops each (i.e. 36 workshops in total) to between 4 and 5 schools each. However, given that one Delivery Centre was discontinued after the first workshop the costs relating to venue hire and delivery staff as presented in Table 41 are calculated based on 11 Delivery Centres providing three workshops each and one Delivery Centre providing one workshop (i.e. 34 workshops in total). Given that the school in the 12th Delivery Centre was re-allocated to another Delivery Centre the costs are based on an assumed 4–5 schools per Delivery Centre (i.e. 50 schools in total) across 11 Delivery Centres where this is used to calculate costs (as opposed to the 3.4 schools per Delivery Centre that in fact occurred) as it better reflects the overall capacity of the workshops, although we are aware that other teachers from other schools also attended.[35] In the first year of the main trial, the following costs were incurred and are all associated with the training:

- Programme Lead Briefing. One day a term during the academic year 2017/2018), three in total, attended by two facilitators from each Delivery Centre, and Subject Matter Experts (SMEs) (33 attendees at the first briefing, 29 at the two subsequent briefings) to assess and facilitate the ongoing programme of workshop delivery. Associated costs included travel, venue hire and refreshments, and printing.
- Workshop Delivery. Three workshops delivered per Delivery Centre for 11 Delivery Centres, and a further workshop by one Delivery Centre (34 in total), co-delivered by SMEs and Delivery Centres and attended by teachers taking part in the evaluation from intervention schools. Costs included Delivery Centre facilitator and SME time, room hire, refreshments, and printing.

[35] These other teachers are not, however, included in the overall costs as they were not participating in the evaluation.

Table 41: Financial costs for 'EAL in the Mainstream Classroom', delivery costs: Main trail (2017/2019)

Item Personnel	Type of cost	Unit cost (£)	Number of units	Total cost per school over three years (£)	Total cost per pupil per year over three years (£)*
Challenge Partners					
Programme management	Set up	65,000.00	1	1,300.00	6.77
Quality assurance costs (including travel)	Set up	800.00	11	176.00	0.92
SMEs**					
Programme Leads Briefing	Set up	1,000.00	3	60.00	0.31
Co-delivery of training workshops by SMEs	Set up	650.00	69	897.00	4.67
School facilitators					
Co-delivery	Set up	350.00	136	952.00	4.96
Travel/subsistence					
Challenge Partners					
Delivery Centre visits	Set up	150.00	33***	99.00	0.52
Challenge Partners and SMEs					
Programme Lead Briefing	Set up	250.00	15	75.00	0.39
Travel, accommodation and subsistence for workshops	Set up	250.00	68	75.00	0.39
Delivery Centres					
Programme Lead Briefing	Set up	250.00	68	340.00	1.77
Venue/refreshments					

Programme Lead Briefings					
Venue hire	Set up	500.00	3	30.00	0.16
Refreshments	Set up	10.00	83	16.60	0.09
Workshops					
Venue hire	Set up	150.00	34	102.00	0.53
Refreshments	Set up	190.00	34	129.20	0.67
Photocopying/printing					
Challenge Partners					
Office costs	Set up	8750.00	1	175.00	0.91
Printing – Programme Lead Briefings	Set up	5.00	83	8.30	0.04
Printing – workshops	Set up	180	34	122.4	0.64
Total				**4,557.60**	**23.74**

[*] *Based on an assumption of 64 Year 10 pupils per school (EAL and FLE). See below for further details.*
[**] *Hounslow Language Service.*
[***] *Assuming 1 DC (out of the 12) not visited.*

School costs

During the interviews, teachers were asked about additional costs for the school for attending training and for implementing the programme. The only additional costs mentioned (outside of staff time) were staff travel to training and printing of worksheets and resources for pupils. The distance between each school and their Delivery Centre was checked and travel costed at 0.23 pence per mile. Given that we have taken travel as occurring from school to Delivery Centre, we have also assumed car share, with one car per school although we recognise that this may not have been the case and travel by tube or train may also have been the case. Given the wide range of distances travelled to Delivery Centres, any travel costs which may be incurred due to attendance at training, should be fully understood by schools prior to committing to three days of workshops. Travel costs are given in Table 42.

As mentioned above, printing was mentioned as a possible source of costs associated with implementation of the programme. However, teachers did not appear to see this as an over-onerous amount above and beyond typical printing costs outside the intervention and counteracted by the increased emphasis on using textbooks. These costs have not, therefore been included in the costing, although schools should be aware of this possibility if they are interested in taking up the intervention.

Table 42: Estimated school costs associated with attendance at training and printing

Item	Type of cost	Cost	Total cost per school over 3 years	Total cost per pupil per year over 3 years
Mileage	Set up	44 miles per school (min. 2; max.146) @0.23p a mile x3 £30.36 per school	£30.36 per school (min £1.38; max. £100.74)	
Printing	Running cost per school	Minimal	-	-
Total			**£30.36**	**0.16**

Time costs

There were three full-days training sessions for participants (up to four teachers per school). This is a considerable time investment by schools and by teachers. Teachers were also expected to complete pre- and intersessional tasks as part of the training. Teachers indicated in the interviews, that they were only able to attend training if there was existing cover within the school. Hence, whilst there were no direct costs assisted with teacher-time (although the sample was limited) there were additional pressures on individual teachers and schools in accommodating attendance at training (and for the teachers in completion of intersessional tasks and lesson planning). Again, schools wishing to take up the training for the programme may, therefore, wish to consider the costs of supply cover for three full-days training.

Number of pupils

In this evaluation, there were on average 64 Year 10 pupils per school in the intervention group (EAL and FLE). As indicated above, the intention was to deliver the programme as a whole-school intervention, which would reduce travel costs (assuming training took place within schools) and increase the number of pupils receiving the programme. This has not, however, been factored into these calculations. The above calculations therefore, assume 192 Year 10 pupils per school over a three-year period.

Table 43 therefore provides the cumulative cost of the programme, including the cost of implementation for a per pupil year for a three-year period. Given that more pupils could, and in high-implementing schools did, receive the programme outside of Year 10 these costs would be considerably reduced, as they would also be if the pilot year was not taken into account.

Table 43: Cumulative costs of EAL in the Mainstream Classroom (assuming implementation over three years)

Programme	Year 1	Year 2	Year 3
EAL in the Mainstream Classroom	23.90	23.90	23.90

Conclusion

Table 44: Key conclusions

Conclusions
EAL pupils in the 'EAL in the Mainstream Classroom' intervention schools made the equivalent of one month's additional progress in GCSE science (GCSE combined or the three separate science GCSEs), on average, compared to EAL pupils in the control group. This result has a moderate to high security rating
EAL pupils in the 'EAL in the Mainstream Classroom' intervention schools also made the equivalent of one month's additional progress in GCSE English language, on average, compared to EAL pupils in the control group. EAL pupils in the 'EAL in the Mainstream Classroom' intervention schools made, on average, no additional progress in GCSE history compared to EAL pupils in the control group
EAL pupils who were eligible for FSM made the equivalent of one month's additional progress in science, on average, compared to EAL pupils eligible for FSM in the control group. However, as the number of EAL pupils eligible for FSM was small, these results should be interpreted with caution
73% of surveyed teachers who received the programme reported they were 'strongly' or 'somewhat' confident in the programme's ability to impact positively on pupils' learning. They also reported increased confidence in supporting EAL pupils, resulting in positive perceptions on EAL pupils' engagement and learning outcomes
The delivery and effectiveness of the programme was likely impacted by delivery challenges including declining workshop attendance, non-completion of assigned tasks, limited support from schools, and a lack of perceived relevance from some teachers

Impact evaluation and IPE integration

Interpretation

The primary outcome analysis shows a small, adjusted effect size of 0.06 suggesting one month's additional progress in GCSE science attainment for those Year 10 EAL pupils that were taught by an 'EAL in the Mainstream Classroom' intervention teacher. Given that the CI includes '0' we cannot be confident that there was a difference between the two groups and the academic progress could have reached from a disadvantage of one month's less progress to an advantage of two months' more progress for those in the intervention group. However, this finding does accord with the positive outcomes of the programme perceived by the 'EAL in the Mainstream Classroom' intervention science teachers who completed the survey at the end of the first year of the evaluation, when nearly three-quarters of respondents (73% or 29 teachers out of 40) agreed (either 'strongly' or 'somewhat') that they were confident in the 'EAL in the Mainstream Classroom' programme's ability to impact positively on pupils learning. At this timepoint they were also more likely to report their Year 10 EAL science pupils were 'very well' engaged in their lessons compared to teachers in the control group and compared to their own reported perceptions of student engagement prior to undertaking the training. Similar findings were also reported by this cohort relating to Year 10 EAL science pupils' understanding of the importance of subject-specific language and confidence in using subject-specific language.

These findings should be placed in the context of a perceived need to improve educational outcomes for EAL pupils within participating schools and the fact that, whilst nearly two-thirds of science teachers (control and intervention) who completed the baseline survey felt that their school supported EAL learners ('very' or 'quite') well, alongside an evident lack of awareness in some cases of what school-level support was available to support EAL pupils, which suggests that this was seen as more of a school-level provision rather than one which encroached greatly in the classroom. In addition, less than half felt ('very' or 'well') able to support EAL learners in their classrooms and less than a third felt confident in doing so. These findings support those of Murphy and Unthiah (2015) relating to a lack of EAL pedagogy in UK schools and the need identified by Strand and Lindorff (2021) for CPD for teachers to support EAL learners. That this programme may go some way towards addressing those issues is reflected in the findings that by the end of the first year of the evaluation nearly three-quarters of 'EAL in the Mainstream Classroom' intervention teachers indicated (in the survey) that they were *confident* in the 'EAL in the Mainstream Classroom' programme's ability to impact positively on pupils learning, they felt better *able* to support EAL pupils in their classroom than their control group peers and then they had prior to taking part in the programme. 'EAL in the Mainstream Classroom' teachers were also more likely to indicate that they felt confident to support EAL pupils in their classroom when compared to teachers in the control group and more

than they did prior to taking the 'EAL in the Mainstream Classroom' training. In intervention schools, taking part in the programme also appeared, in the short term at least, to raise awareness of in-school provision to support EAL learners.

The CACE analysis does not show any differential effect based on our compliance measure, attendance at training events. Descriptively, levels of attendance at the programme CPD workshops were variable: Only 60% of intervention science teachers attended all three training workshops, 21% attended two, 11% only attended one (predominantly the first workshop), and a further 9% did not attend any workshops. And at the same time as attendance declined during the six-month period over workshop delivery, the value placed on the workshops appeared to increase among those continuing to attend. Workshops 2 and 3 were perceived as being more user-friendly in terms of providing useful strategies and resources of direct relevance to teachers' classroom practice (compared to the perceived grammar-heavy content of Workshop 1), although some teachers did feel that more time could have been spent in the CPD on planning and creating their own resources. This suggests that there may be a case for changing the structure of the workshop content distribution to maintain teacher engagement.

It is likely that only those science teachers who were most engaged with the programme attended (or continued to attend) the CPD. However, there appear to have been a number of other factors also at work; namely, competing pressures on teachers' time and teachers' perceptions of the value of the training. School buy-in is an important factor for any CPD at this level. Overall, however, there appears that teachers lacked awareness of the programme's requirements (in terms of the types of training, especially language learning) and the amount of time needed not only to attend training but also to develop and tailor strategies to meet their own pupils' needs, and to complete the intersessional tasks. The need to send teachers for three full-day's training was problematic for schools, although they were spread out over two terms. Underlying this, there also appears to have been a lack of school support for the programme overall with no earmarking of attendance at training over other school workload and the training was competing with wider capacity issues. This was compounded by a lack of support from the programme between sessions and a lack of monitoring of completion of intersessional tasks. In addition, time constraints, which meant that teachers were unable to attend training led to an acceptance on the part of the delivery team of cascading occurring. Although in the event this did not occur to a large degree this would, perhaps, inevitably, have led to dilution of the programme training for the small number of teachers who reported receiving cascaded training from colleagues (n = 5) particularly as there were no formal mechanisms in place for this process.

In addition, the majority of science teacher survey respondents (84%) indicated that they felt that the CPD provided by the workshops overall was ('very' or 'somewhat') effective and teachers valued the chance to better understand the needs of their EAL pupils and learning strategies that they could see were directly transferable to their classroom context. However, the CACE analysis indicated that there was no differential effect on outcomes based on whether or not intervention teachers attended any training or between the amount of training attended. It may be that enrolling on the programme raised awareness in teachers of the importance of teaching academic language (in terms of both vocabulary and grammar) leading to perceptions by teachers of a positive impact of the CPD. This may also be supported by the increase in awareness of in-school support for EAL pupils seen in the survey findings for the 'EAL in the Mainstream Classroom' intervention teachers at the end of the first year of the evaluation (Figure 7, above). In addition, the survey findings provide some support for the existing evidence (Strand and Lindorff, 2021) that there is varied understanding by teachers of EAL pupils' proficiency in English (given that the role of the programme in enabling teachers to better understand the needs of their EAL pupils was particularly welcomed by teachers; see section 'Teachers learning'), which the CPD may well have ameliorated.

The CPD was spread out over two terms to allow for teachers to gradually embed the programme and develop over time. This aspect did appear to work well, ensuring that teachers who attended training did maintain an awareness of the programme and programme strategies during the first main trial year. The intersessional tasks did prompt them to continue implementation, even if at a low level, between sessions. The majority of intervention survey respondents agreed (either 'strongly' or 'somewhat') that they felt confident in their ability to deliver the 'EAL in the Mainstream Classroom', that the programme had changed their pedagogic practice and that they felt that they had been able to implement the learning from the programme either 'very' or 'somewhat' effectively at the end of the first trial year (and two school terms). Finding time to make changes to pedagogic practice was a challenge for teachers, compounded by the introduction of a new science curriculum in Key Stage 4. In addition, many science teachers who failed to see the relevance of the more grammatical content of the CPD indicated that they did not feel that the programme was well-suited to their subject (science). This view was particularly pertinent as the new Key Stage 4 science curriculum was not perceived (by teachers) to place a high value on writing skills. Furthermore, only two-thirds of the 'EAL in the

Mainstream Classroom' survey respondents indicated that they were teaching the same class(es) in the second year of the evaluation (when the pupils were in Year 11) and less than a third of these indicated that they continued to use learning from the programme 'to a great extent'. Consequently, the results should be interpreted within the context of variable implementation within a short period of time. This lack of prolonged (and variable levels of) implementation can be deduced to have had reduced any potentially positive outcomes of as it was not delivered as intended (i.e. over a sustained period of time). It also suggests that a longer period of support or engagement with the programme may be necessary to potentially show positive effects with a greater degree of statistical certainty and aligns with arguments for interventions to be embedded for a longer period of time than that in this trial (i.e. 2 years+) (Matuchniak *et al.*, 2014; Maerten-Rivera *et al.*, 2016) and possibly supporting the significance of ongoing CPD (Lara-Alecio *et al.*, 2012).

In terms of the secondary research question, the analysis of how effective the 'EAL in the Mainstream Classroom' programme was in improving subject-specific academic attainment in a second GCSE subject (history), again showed higher average scores for EAL pupils in the intervention group than for those in the control group. However, the adjusted effect size based on the analytic model is small (0.04), which in EEF terms equates to no additional month's progress. As before, the CI includes 0, which means that we cannot be confident that there was a difference between the two groups and the academic progress could have reached from a disadvantage of two months' less progress to an advantage of two months' more progress for those in the intervention group. This appears at first surprising, given that history is typically seen as a more context embedded subject and therefore perhaps perceived as more appropriate a subject within which to assess the impact of the programme. It may be that much of the vocabulary and grammar-teaching was already occurring within history classroom practice (although given that the IPE focused on science teachers, we do not know this, see 'Limitations' section below). Or it may be that with a longer time period to embed the programme effects may have been seen on pupils' attainment in this subject. Alternatively, it may be that the technical language in science was particularly amenable to the strategies taught within the programme (e.g. nominalisation).

Either way, these findings must be interpreted alongside the findings for the third research question: how effective the 'EAL in the Mainstream Classroom' programme is in improving academic attainment in English (as measured by GCSE English language) when delivered to Key Stage 4 EAL pupils? This was chosen to assess any indirect effects through the wider curriculum in EAL pupils' learning given that English language teachers were not part of the 'EAL in the Mainstream Classroom' CPD but, like history, English language is a context embedded subject (arguably even more so). Again, EAL pupils taught by 'EAL in the Mainstream Classroom' intervention teachers showed higher scores on average than those pupils in the control group, with an adjusted effect size of 0.07, which equates to one month's additional progress. Again, however, the CI includes '0', which means that we cannot be confident that there was a difference between the two groups and the academic progress could have reached from no additional progress to an advantage of two months' more progress.

Whilst the analysis reported above was looking purely at diffusion effects, a subsequent analysis was looking for potential cumulative benefits from the intervention. This analysis aimed to assess the impact of 'EAL in the Mainstream Classroom' when pupils receive the approach from more than one teacher in more than one subject area (i.e. when pupils are taught by 'EAL in the Mainstream Classroom' teachers in both science and history GCSE subjects). Again, using GCSE English language attainment as the outcome measure, we observed a weak relationship in the expected direction, namely pupils with teachers in both disciplines showed on average 0.17 higher GCSE English language results than pupils in the control group (effect size of 0.12, which equates to two months' additional progress). As before, the coefficient's CI included '0' and we therefore concluded that there is insufficient evidence to interpret these findings with confidence. Consequently, there is only limited evidence that either diffusion of the programme or cumulative gains occurred in the time period of the evaluation.

The analysis relating to the impact of 'EAL in the Mainstream Classroom' on non-EAL pupils within the same classrooms, again using GCSE English language attainment as the outcome measure found no evidence of a difference in achievement for EAL pupils when compared to their non-EAL peers (i.e. the programme did not favour one group over another). Findings from the science teacher survey indicated that 'EAL in the Mainstream Classroom' intervention teachers were more likely to report positive perceptions of non-EAL pupils than they were for their EAL counterparts in their Year 10 science classrooms. They were more likely to also report that their FLE pupils engaged 'very well' during their lessons than teachers in the control group (26% or 11 intervention teachers compared to 16% or 5 teachers), and were more likely to report that their FLE pupils understood the importance of subject-specific language 'very' or 'quite' well (91% or 38 intervention teachers compared to 77% or 24 control group teachers). In contrast, teachers in the control group (77% or 24 teachers) rated their FLE pupils as 'very' or 'quite' confident using subject-specific language compared

to 67% (or 28 teachers) of the intervention teachers. However, overall, the numbers included in the survey sample are small and we do not wish to over-emphasise their importance, rather raise them as a possible trend.

There were no differential effects found on the subgroup analysis of those pupils in receipt of FSM or for those with differing proficiency in English bands suggesting no one subgroup was advantaged or disadvantaged as a result of receiving the intervention. As statistical power for such analyses was low, interested readers are directed to Appendix O for details on estimates and their associated CIs.

The final subgroup analysis tested whether EAL students' choice regarding combined science/three separate science GCSEs introduced heterogeneity in the programme's effect due to possibly different profiles of pupils taking these options (students with higher prior attainment being more likely than their peers to take the three separate GCSEs in biology, chemistry, and physics). As expected, pupils who selected the three separate science subjects had on average a 1.13 higher GCSE science score than those who did not (CI did not include '0'). However, there was no cross-level interaction effect with treatment, i.e. pupils in either GCSE science option were unlikely to differentially benefit from the intervention.

Our overall conclusions are that the intervention was not evaluated as having shown a potential effect. Descriptively, the intervention was associated with one month's additional progress in science on student outcomes, but due to the statistical uncertainty around that result the academic progress could have reached from a disadvantage of one months' less progress to an advantage of two months' more progress for those in the intervention group. Additionally, none of the estimated effect sizes for the primary and secondary analyses are in the range that was originally seen as potentially relevant for classroom practice (i.e. MDES ≥ |0.20|; see Statistical Analysis Plan). And these conclusions remain the same under planned additional analyses. Whilst the IPE suggests some changes were occurring within schools as a result of the programme the number or participants, for example completing the survey or taking part in the interviews/observations, are small meaning we should be cautious in our interpretation. However, overall this evaluation does in some way add to the current limited evidence base on potential interventions to support the academic attainment of EAL pupils, particularly previous research on ongoing workshops to support teachers in integrating literacy and language into the curriculum (Lara-Alecio *et al.*, 2012) whilst it suggests the need for a longer period of embedding of the CPD within the classroom in order to be able to truly detect an impact (if any).

Evidence to support the logic model

Whilst the impact evaluation analysis urges caution in interpreting overall findings, some aspects of the logic model did appear to work well in some cases. Generally, teachers were very positive about learning about the needs of their EAL pupils and the ways in which the programme enabled them to develop their understanding of the student in a more holistic way. Others did appreciate the language aspects of the CPD and, as indicated above, felt that many of the strategies of the programme were directly relevant and easy to implement within their own classroom setting. Teachers agreed with the underlying principles of the programme and agreed that it was a potential vehicle for change, *if* they had sufficient time to implement in order to make an impact. Most importantly, they indicated that they had changed their pedagogic practice as a result of the programme, although this is purely based on self-report.

Some earlier mechanisms within the casual chain were, however, weak. Monitoring by Challenge Partners and the Delivery Centres was variable, as was attendance at training. This meant there was insufficient evidence to gauge the completion of intersessional tasks. When teachers had the time to do the tasks, they did find them useful, and they did appear to encourage implementation. There is some limited evidence in the IPE, however, that this lack of monitoring and support did deter participants from completion (see section 'Experience of training'), which meant that subsequent implementation was therefore also impacted. Time constraints felt by teachers and lack of attendance at training (or subsequent cascading) can only be concluded to have meant that the programme was not delivered to the extent intended across all intervention schools.

It is also important to note that the logic model is not explicit about the mechanisms for change underlying the programme i.e. the ways in which in teachers' practices should change as a result of an increased understanding of EAL learners' needs. It is this linking of knowledge and practice, which has also be found to be a key weakness of implementation in this evaluation. This was compounded by a lack of formal mechanisms for cascading training, monitoring, and support for completion of intersessional tasks and teachers' perceived lack of time to plan strategies in the training sessions.

Finally, there was no evidence that emerging leaders were targeted to participate in the CPD. Although some cascading of the approaches learned within the training was reported to occur within intervention schools, there was little cascading to intervention teachers unable to attend training, even though this was provided for within the programme.[36] It was posited [in the Intervention Delivery and Evaluation Analysis (IDEA) workshop] that by targeting emerging leaders, a more 'whole-school' approach could be taken to language and literacy teaching within the mainstream classroom and thereby provide greater levels of exposure to the programme by pupils. However, these were more tentative aspects of the logic model and thereby not too much weight should be placed on their absence.

Limitations and lessons learned

This was a two-armed school-level randomised controlled trial. The programme was designed to meet a perceived need in the recruited schools and addressed an area (the needs of EAL pupils within the mainstream classroom) teachers felt unsure about. However, recruitment was an issue and lower than expected, which had an overall impact on the security of the findings.

Changes to the recruitment plan (i.e. the removal of Cohort II) and the finding that ICCs were much higher than originally expected (Table 9) all contributed to the reduced statistical power of the study. Consequently, at randomisation and analysis stages only effect sizes of medium size could have been detected via significance tests (MDES = 0.31 and MDES = 0.44, respectively; Cohen, 1988; Spybrook *et al.,* 2016), compared to the originally planned effect size of MDES = .20. Nevertheless, it should be noted that the primary and secondary analyses resulted in 95% CIs around '0' for the effect size estimates (see Table 12; e.g. for the primary outcome the CI ranged from -0.06 to 0.18), which did not include effect sizes that were originally judged as potentially relevant for practice (i.e., MDES = 0.20 or larger in favour of the intervention group), indicating, based on the current data and analyses that, the expected impact, if indeed present, might be smaller than this threshold. It is also important to note that more recently, meta research is re-evaluating the characterisation of MDES = 0.20 as a 'small' effect size and questions whether trials planned to this standard are sufficiently sensitive to detect differences that are relevant for school contexts, especially when evaluating student achievement (Kraft, 2020).

The changes to the recruitment plan also had consequences for the allocation of schools to the intervention groups. As the schools included in the trial (originally Cohort I) were randomised in the last quarter of 2017 and recruitment was both stratified by Delivery Centre and occurred in batches, the number of schools per group were less equal than one would usually expect with either a larger number of schools (in this case including Cohort II, plans for which were discontinued in October 2018) or without stratification or batches (see discussion in Grischott, 2018). This way the sum of randomly occurring unequal allocations per Delivery Centre led to more schools being allocated to the intervention group than to the control group. We additionally investigated the potential contribution of Delivery Centres in terms of explained variance in science scores (i.e. potential alternative explanation for the analysis result) and the Key Stage 2 reading scores (i.e. potential selection effects). The result showed that less than 1% of the variance in either variable could be attributed to them, i.e. their potential impact on the result of the primary outcome was minimal (see Appendix O, Table 13).

Both the impact and process evaluation identified imbalances in the two samples at baseline regarding school characteristics, as well as teacher and pupil composition. For example, at the school level, intervention schools had higher proportions of pupils with EAL and higher proportions of EAL pupils achieving strong GCSE passes in English and maths, which would suggest that they already had a higher 'value-added' component within their school structure, and therefore may mean that pre-existing advantages in the baseline variables may be reflected in the outcome (i.e. the findings may be a result of already existing school differences).

Some of the samples analysed for the secondary questions were imbalanced by the baseline measure of the Key Stage 2 SATs reading attainment (higher average baseline scores in the control group), EAL proficiency in English (higher proficiency in English bands in the control group), and FSM (lower share in the control group). Analyses including these variables to evaluate the sensitivity of results to these imbalances nevertheless showed the same results as the planned analyses. Whilst the sample for the primary analysis was balanced with view to these potential confounders, presumably because science is compulsory at Key Stage 4, especially the sample for history (which is an optional GCSE subject)

[36] Whilst this was not included in the logic model it was subsequently allowed as a result of individual requests by schools.

would have likely benefited from a stratification approach at the school level (which was discussed but deemed impossible before recruitment and minimisation).

Although on average the recruitment was as expected or even above the planned numbers in the Statistical Analysis Plan (Table 10), a further limitation needs to be kept in mind when assessing the statistical analysis: in the primary and secondary outcome analyses every school provided at least one student, but some schools (exact number not disclosed for SDC) contributed with only one student to the analyses. This was unexpected, since the evaluation team and Delivery Centres ensured that schools meet the eligibility criteria relating to the minimum required number of teachers and EAL pupils with Key Stage 2 results (as reported by the schools). This information became only available after pupil data was matched with the NPD data. EAL pupils had a probability of 0.029 that their pupil-level data was unable to be matched in the NPD; non-EAL pupils a probability of 0.020 (odds ratio [OR] = 1.49; SE = 0.19; 95% CI: from 1.03 to 2.16). Therefore, differential attrition due to our inability to match some EAL students with NPD records was observed. EAL pupils in schools allocated to the intervention condition had a probability of 0.027 to be unmatched; and EAL pupils in schools allocated to the control condition had a probability of 0.022 to be unmatched (OR = 1.26, SE = 0.18; 95% CI: from 0.89 to 1.79). Overall, effect sizes as well as the relative numbers were nevertheless small and with limited impact on the findings of this trial. The main consequence was that the often low numbers of EAL pupils per school made it impossible to estimate random effects for slopes in the moderation analyses, which is conceptually preferable when there is a potential for school-specific effects of interventions (e.g. Preacher *et al.*, 2010).

Whilst the IPE was thorough, the number of case studies was impacted by school withdrawal through the requirement to sign the GDPR-compliant addendum. The numbers of responses to the survey were in some cases small, with only 63% of all participating teachers completing the survey at T2 and a comparable 63% of all intervention teachers completing the survey at T3,[37] which may have introduced some selection bias into the survey findings. However, the open-ended responses within the survey and the interviews did provide rich data. This means that the findings from the surveys should be viewed with some caution. Whilst the lesson observations were conducted by the researchers who had attended the training, the nature of the programme means that strategies were designed to be adapted, which may have impacted on the findings from that section of IPE. Variability in the collection of routine monitoring data also had implications for the interpretation of findings, particularly relating to compliance, training, and intersessional tasks. In addition, implementation appears to have taken time to embed, so a longer period of the evaluation, more, less intensive training sessions giving teachers' time to discuss and reflect and increased programme support between the training days may have been beneficial.

The impact evaluation was designed to only include those pupils who were in the English education system at the end of Key Stage 2. This was pragmatic, due to the need for a baseline measure without overburdening schools, and in line with the theory of change of the programme. Whilst not included in the logic model, it was implicit within the programme that the strategies would be most beneficial to pupils at proficiency levels C and D. Teachers also felt that lower attaining pupils benefited most (by which we perceive proficiency bands C and D given that level A pupils are generally provided for separately within schools and the difficulty of fully accessing the curriculum at those proficiency levels). In this evaluation, pupils were drawn from across the prior attainment levels although some teachers of higher prior attaining classes also perceived benefits for their pupils.

Research suggests that it is principally pupils below proficiency band C who significantly underachieve at GCSE. As seen in the impact evaluation, the substantial majority of EAL pupils included already had good proficiency levels at or above the programme's goals (levels C and D) meaning this aspect of the programme was not sufficiently testable. As the impact evaluation excluded pupils that had not been within the English education system at the end of Key Stage 2 it is likely that some EAL pupils at lower proficiency levels who may have benefited from the programme were not included in the analyses. In addition, the use of Key Stage 2 assessments systematically underestimates GCSE achievement for pupils that had low proficiency in English at Key Stage 2 due to accelerated attainment as proficiency increases through secondary school (Strand *et al.*, 2015). Consequently, these factors may have skewed the overall results of the evaluation given that pupils' proficiency at the time of Key Stage 2 assessments was not taken into account at the analysis stage (and this data was not mandatorily collected by schools at the time) and our sample predominantly consisted of those pupils with higher levels of proficiency in English in Key Stage 4. Finally, the recruited schools had a

[37] The T3 survey was only administered to teachers allocated to the intervention group

high proportion of pupils for whom was EAL (49% of intervention school pupils were classed as EAL), which would mean that the findings should not really be generalised to schools with low levels of EAL pupils.

Lack of school buy-in in terms of teacher-time impacted on both attendance at training and implementation. Workshop attendance declined after the first workshop, which may have been because teachers struggled with the 'academic language' aspects of the programme (and at this point three schools withdrew from implementing the programme). In addition, some science teachers did not see the relevance of language and literacy teaching to the Key Stage 4 science curriculum (and in some cases, science more generally). Due to these factors, it was beneficial to include history as a second subject, given that the content of the programme could perhaps be argued to be perceived by teachers as of more direct relevance. However, given that the IPE focused on science teachers we do not know any more on the mechanism for change within that subject, or different subjects more generally, which perhaps a whole-school approach would have facilitated. It should also be noted that at its most effective, implementation in this trial only occurred over five academic terms. If the 'EAL in the Mainstream Classroom' programme was to be re-evaluated we would recommend a whole-school approach, across subjects and starting in Key Stage 3.

A final aspect of learning from this project is that we present estimates for how well Key Stage 2 data from the NPD can be used as baseline data for studies researching questions involving EAL pupils in post-Key Stage 2 education. Figures 3 and 4 (CONSORT) show that EAL pupils were less likely to have Key Stage 2 reading scores documented in the NPD. Looking at only those pupils who could be matched with NPD records (see section on NPD matching above), EAL pupils had a probability of 0.203 to have no documented Key Stage 2 reading score; whilst non-EAL pupils had a probability of 0.036 (OR = 6.86, SE = 0.12; 95% CI: from 5.42 to 8.68), i.e. EAL pupils were substantially more likely to have no documented Key Stage 2 reading scores. As discussed in other places of this report, this was anticipated by all stakeholders and contributed to the decision to have Key Stage 2 results as an eligibility criterion. But there is also an additional component to this, as this is another variable that needs to be considered when setting up allocation procedures. Our analyses showed that for those pupils who could be matched with NPD records, pupils in schools allocated to the intervention condition had a probability of 0.14 to have no documented Key Stage 2 reading score; and pupils in schools allocated to the control condition had a probability of 0.08 to have no documented Key Stage 2 reading score (OR = 1.83, SE = 0.09; 95% CI: from 1.53 to 2.13). This effect size points to a relevant difference between the two conditions, which in our case would have had limited impact as Key Stage 2 results were used as a baseline control variable on an individual level to control for eventual differences. Nevertheless, whilst there are distinct advantages from a power perspective to include baseline control variables (as recognised by the EEF guidelines for trial planning) as well as efficiency gains if the data are available in the NPD, our results highlight that from a representation and inclusion perspective, studies investigating interventions that target all EAL pupils in post-Key Stage 2 education might explore the potential of different designs, such as using endpoint-only designs or to conduct their own baseline proficiency assessment to be able to consider all EAL pupils.[38]

Future research and publications

This evaluation raised some important issues relating to teachers' understanding of the needs of EAL pupils and mechanisms for support. One area of this report, which the research team would like to publish on further, relates to teachers' understanding and awareness of support for EAL pupils. Given the positive outcomes of the programme as perceived by the participating teachers the mechanisms for change to enable these to translate into improved educational outcomes for EAL (and FLE) pupils also would be an interesting area of inquiry. Given the exploratory nature of much of the analysis in this report the programme itself may benefit from further research into the model as a whole-school approach with a revised logic model identifying evidence-informed mechanisms for changes in teachers' practice and, perhaps, a longer period of implementation and delivery support than occurred within this evaluation.

[38] This may also have improved recruitment to this trial as some schools struggled to have sufficient numbers of pupils to meet the evaluation eligibility criteria.

References

August, D., Branum-Martin, L., Cárdenas-Hagan, E., Francis, D.J., Powell, J., Moore, S. and Haynes, E.F. (2014). 'Helping ELLs Meet the Common Core State Standards for Literacy in Science: The Impact of an Instructional Intervention Focused on Academic Language'. *Journal of Research on Educational Effectiveness*, 7: 1, 54–82. https://doi.org/10.1080/19345747.2013.836763

Bates, D., Mächler, M., Bolker, B. and Walker, S. (2015). 'Fitting Linear Mixed-Effects Models Using lme4'. *Journal of Statistical Software*, 67: 1, 1–48. https://doi.org/10.18637/jss.v067.i01

Benton, T. and Sutch, T. (2013). 'Exploring the Value of GCSE Prediction Matrices Based Upon Attainment at Key Stage 2. Cambridge Assessment Research Report'. Cambridge, UK: Cambridge Assessment. Available at: https://www.cambridgeassessment.org.uk/Images/181034-exploring-the-value-of-gcse-prediction-matrices-based-upon-attainment-at-key-stage-2.pdf (accessed 20 June 2022).

Bingham, A.J. and Witkowsky, P. (2022). 'Deductive and Inductive Approaches to Qualitative Data Analysis'. In C. Vanover, P. Mihas, and J. Saldaña (Eds.), 'Analyzing and Interpreting Qualitative Data: After the Interview' (pp. 133–46). London, UK: SAGE Publications, Inc.

Bland, J.M. and Altman, D.G. (2015). 'Statistics Notes: Bootstrap Resampling Methods'. *British Medical Journal*, 350: h2622. https://doi.org/10.1136/bmj.h2622

Carroll, J.E. (2016). 'The Whole Point of the Thing: How Nominalisation Might Develop Students' Written Causal Arguments'. *Teaching History,* 162: 16–24.

Challenge Partners. (2017a). 'Delegate Pre-Programme Handbook'. London, UK: Challenge Partners.

Challenge Partners. (2017b). 'EAL in the Mainstream Classroom. Delivery Centre Handbook'. London, UK: Challenge Partners.

Charlton, C., Rasbash, J., Browne, W.J., Healy, M. and Cameron, B. (2019). 'MLwiN Version 3.03'. Bristol, UK: Centre for Multilevel Modelling, University of Bristol.

Cohen, J. (1988). 'Statistical Power Analysis for the Behavioral Sciences (2nd ed.)'. New Jersey, US: Lawrence Erlbaum Associates.

Crosson, A.C. and Moore, D. (2017). 'When to Take Up Roots: The Effects of Morphology Instruction for Middle School and High School English Learners'. *Reading Psychology,* 38: 3, 262–88. https://doi.org/10.1080/02702711.2016.1263699

Dare, B. (2010). 'Learning About Language: The Role of Metalanguage'. *NALDIC Quarterly,* 8: 1, 18–25. http://oro.open.ac.uk/25026/1/NaldicSFL.pdf

Department for Education (DfE). (2018). 'School Census 2017 to 2018 Guide, Version 1.8. May 2018'. London, UK: DfE. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/708550/2017_to_2018_School_Census_Guide_V1_8.pdf (accessed 20 June 2022).

Department for Education (DfE). (2020). 'English Proficiency of Pupils with English as an Additional Language. Ad-Hoc Notice. February 2020'. London, UK: DfE. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/868209/English_proficiency_of_EAL_pupils.pdf (accessed 20 June 2022).

Department for Education (DfE). (2021). 'School pupils and their characteristics. Academic year 2020/2021'. Available at: **https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics/2020-21** (accessed 18 June 2022).

Education Endowment Foundation (EEF). (2013). 'Pre-Testing in EEF Evaluations'. London, UK: EEF.

Education Endowment Foundation (EEF). (2016). 'Cost Evaluation Guidance for EEF Evaluations'. London, UK: EEF.

Education Endowment Foundation (EEF). (2018). 'Statistical Analysis Guidance for EEF Evaluations'. London, UK: EEF.

Education Endowment Foundation (EEF). (2019). 'Cost Evaluation Guidance for EEF Evaluations'. London, UK: EEF.

Enders, C.K., Keller, B.T. and Levy, R. (2018). 'A Fully Conditional Specification Approach to Multilevel Imputation of Categorical and Continuous Variables'. *Psychological Methods,* 23: 2, 298–317. https://psycnet.apa.org/doi/10.1037/met0000148

Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). 'G*Power 3: A Flexible Statistical Power Analysis Program for the Social, Behavioral, and Biomedical Sciences'. *Behavior Research Methods*, 39: 2, 175–91. https://doi.org/10.3758/BF03193146

Gardner, M.J. and Altman, D.G. (1988). 'Estimating with Confidence'. *British Medical Journal*, 296, 1210–11. https://doi.org/10.1136/bmj.296.6631.1210

Graham, L.J., White, S.L.J., Cologon, K. and Pianta, R.C. (2020) 'Do Teachers' Years of Experience Make a Difference in the Quality of Teaching?' *Teaching and Teacher Education*, 96, 103190. https://doi.org/10.1016/j.tate.2020.103190

Greenland, S., Senn, S.J., Rothman, K.J., Carlin, J.B., Poole, C., Goodman, S.N. and Altman, D.G. (2016). 'Statistical Tests, P Values, Confidence Intervals, and Power: A Guide to Misinterpretations'. *European Journal of Epidemiology,* 31: 337–50. https://doi.org/10.1007/s10654-016-0149-3

Grischott, T. (2018). 'The Shiny Balancer—Software and Imbalance Criteria for Optimally Balanced Treatment Allocation in Small RCTs and cRCTs'. *BMC Medical Research Methodology*, 18: 1, 108. https://doi.org/10.1186/s12874-018-0551-5

Hanley,P., Böhnke, J.R., Slavin, R.S., Elliott, L. and Croudace, T. (2016). 'Let's Think Secondary Science. Evaluation Report and Executive Summary'. London, UK: EEF.

Hedges, L.V. (2007). 'Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics*, 32: 4, 341 - 370 https://doi.org/10.3102/1076998606298043

Hollingworth, S. and Mansaray, A. (2012). 'Language Diversity and Attainment in English Secondary Schools: A Scoping Study'. London, UK: The Institute for Policy Studies in Education (IPSE).

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R. and Kerr, K. (2019) Implementation and process evaluation (IPE) for interventions in education settings: An introductory handbook. Education Endowment Foundation. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-design/IPE_Handbook.pdf?v=1708673677 (accessed 23 February 2024)

Hutchinson, J. (2018). 'Educational Outcomes of Children with English as an Additional Language'. Cambridge, UK: The Bell Foundation. Available at: https://www.bell-foundation.org.uk/research-report/educational-outcomes-of-children[1]with-english-as-an-additional-language/ (accessed 1 March 2018).

Hwang, J.K., Lawrence, J,F., Mo, E. and Snow, C. (2015). 'Differential Effects of a Systematic Vocabulary Intervention on Adolescent Language Minority Students With Varying Levels of English Proficiency'. *International Journal of Bilingualism*, 19: 3, 314–32. **https://escholarship.org/uc/item/5wh435nv**

Huang, F.L. (2018). 'Using Cluster Bootstrapping to Analyze Nested Data With a Few Clusters'. *Educational and Psychological Measurement,* 78: 2, 297–318. https://doi.org/10.1177/0013164416678980

Inglis, G., Archibald, D., Doi, L., Laird, Y., Malden, S., Marryat, L., McAteer, J., Pringle, J. and Frank, J. (2018). 'Credibility of Subgroup Analyses by Socioeconomic Status in Public Health Intervention Evaluations: An Underappreciated Problem?' *SSM – Population Health*, 6, 245–51. https://doi.org/10.1016/j.ssmph.2018.09.010

Kraft, M.A. (2020). 'Interpreting Effect Sizes of Education Interventions'. *Educational Researcher,* 49: 4, 241–53. https://doi.org/10.3102/0013189X20912798

Lara-Alecio, R., Tong, F., Irby, B.J., Guerrero, C., Huerta, M. and Fan, Y. (2012). 'The Effect of an Instructional Intervention on Middle School English Learners' Science and English Reading Achievement'. *Journal of Research in Science Teaching,* 49: 8, 987–1011. doi:10.1002/tea.21031

Lumivero (2010) NVivo (Version 9). Available at: www.lumivero.com

Maerten-Rivera, J., Ahn, S., Lanier, K., Diaz, J. and Lee, O. (2016). 'Effect of a Multiyear Intervention on Science Achievement of All Students Including English Language Learners'. *The Elementary School Journal,* 116: 4, 600–24. https://doi.org/10.1086/686250.

Matuchniak, T., Olson, C.B. and R. Scarcella, R. (2014). 'Examining the Text-Based, On-Demand, Analytical Writing of Main-Streamed Latino English Learners in a Randomized Field Trial of the Pathway Project Intervention'. *Reading and Writing*, 27: 6, 973–94. https://doi.org/10.1007/s11145-013-9490-z

Murphy, A. and Unthiah, A. (2015). 'A Systematic Review of Intervention Research Examining English Language and Literacy Development in Children with English as an Additional Language (EAL)'. London, UK: EEF.

NALDIC. (2018). 'The Attainment of EAL Pupils in England – What the Headlines Don't Tell Us'. Available at: **https://naldic.org.uk/the-attainment-of-eal-pupils-in-england-what-the-headlines-dont-tell-us/** (accessed 6 April 2022).

Oxley, E., and de Cat, C. (2021). 'A Systematic Review of Language and Literacy Interventions in Children and Adolescents with English as an Additional Language (EAL)'. *The Language and Learning Journal,* 49: 3, 265–87. https://doi.org/10.1080/09571736.2019.1597146

Petticrew, M., Tugwell, P., Kristjansson, E., Oliver, S., Ueffing, E. and Welch, V. (2012). 'Damned If You Do, Damned If You Don't: Subgroup Analysis and Equity'. *Journal of Epidemiology and Community Health,* 66: 1, 95–8. https://doi.org/10.1136/jech.2010.121095

Pituch, K.A., Stapleton, L.M. and Kang, J.Y. (2006). 'A Comparison of Single Sample and Bootstrap Methods to Assess Mediation in Cluster Randomized Trials'. *Multivariate Behavioral Research,* 41: 3, 367–400. https://doi.org/10.1207/s15327906mbr4103_5

Polias, J. (2010). 'Pedagogical Resonance: Improving Teaching and Learning'. *NALDIC Quarterly,* 8: 1, 42–49. **http://oro.open.ac.uk/25026/1/NaldicSFL.pdf**

Preacher, K.J., Zyphur, M.J. and Zhang, Z. (2010). 'A General Multilevel SEM Framework for Assessing Multilevel Mediation. *Psychological Methods,* 15 : 3, 209–33. **https://doi.org/10.1037/a0020141**

Qualtrics (2015). Qualtrics XM [Software]. Provo, Utah. Available at: https://www.qualtrics.com/

Raudenbush, S.W., Spybrook, J., Bloom, H., Congdon, R., Hill, C. and Martínez, A. (2011). 'Optimal Design. Software for Multi-Level and Longitudinal Research (Version 3.01)'. [Software]. Available at: **www.wtgrantfoundation.org** (accessed 30 September 2022).

R Core Team. (2020). 'R: A Language and Environment for Statistical Computing (Version 4.0.2). Vienna: R Foundation for Statistical Computing'.

Saghaei, M. and Saghaei, S. (2011). 'Implementation of an Open-Source Customizable Minimization Program for Allocation of Patients to Parallel Groups in Clinical Trials'. *Journal of Biomedical Science and Engineering*, 4: 11, 734–39. http://dx.doi.org/10.4236/jbise.2011.411090

Schleppegrell, M.J. (2010) 'Supporting a "Reading to Write" Pedagogy with Functional Grammar'. *NALDIC Quarterly,* 8: 1, 26–30. http://oro.open.ac.uk/25026/1/NaldicSFL.pdf

Sheddon, S. (2015). 'London Schools Excellence Fund. Self-Evaluation Toolkit. Final Report'. London, UK: London West Alliance.

Spybrook, J., Shi, R. and Kelcey, B. (2016). 'Progress in the Past Decade: An Examination of the Precision of Cluster Randomized Trials Funded by the U.S. Institute of Education Sciences'. *International Journal of Research and Method in Education,* 39: 3, 255–67. https://doi.org/10.1080/1743727X.2016.1150454

Steele, F., Vignoles, A. and Jenkins, A. (2007). 'The Effect of School Resources on Pupil Attainment: A Multilevel Simultaneously Equation Modelling Approach'. *Journal of the Royal Statistical Society, Series A,* 170: 3, 801–24. http://www.jstor.org/stable/4623201

Strand, S., Malmberg, L. and Hall, J. (2015). 'English as an Additional Language (EAL) and Educational Achievement in England: An Analysis of the National Pupil Database'. London, UK: EEF.

Strand, S. (2016). 'Educational Outcomes Among Children with English as an Additional Language (EAL)'. Briefings. The Migration Observatory At The University of Oxford. Available at: https://migrationobservatory.ox.ac.uk/resources/briefings/educational-outcomes-among-children-english-additional-language-eal/ (accessed 20 June 2022).

Strand, S. and Lindorff, A. (2021). 'English as an Additional Language, Proficiency in English and Rate of Progression: Pupil, School and LA Variation'. Cambridge, UK: University of Oxford, Unbound Philanthropy, and The Bell Foundation. Available at: https://www.bell-foundation.org.uk/app/uploads/2021/03/University-of-Oxford-Report-March-2021.pdf (accessed 20 June 2022).

The Bell Foundation. 'Great Idea: Dictogloss'. Available at: https://www.bell-foundation.org.uk/eal-programme/guidance/classroom-guidance/great-ideas/dictogloss/#:~:text=Dictogloss%20is%20a%20type%20of,but%20with%20familiar%20subject%20content (accessed 20 June 2022).

Wong Fillmore, L. and Snow, C.E. (2000). 'What Teachers Need to Know About Language'. Clearinghouse on Languages and Linguistics. Special Report. Berkeley, CA: University of California. Available at: https://people.ucsc.edu/~ktellez/wong-fill-snow.html (accessed 20 June 2022).

Xiao, Z., Kasim, A., and Higgins, S.E. (2016). 'Same Difference? Understanding Variation in the Estimation of Effect Sizes from Educational Trials'. *International Journal of Educational Research*, 77: 1-14 http://dx.doi.org/10.1016/j.ijer.2016.02.001

Appendix A: EEF cost rating

Figure 32: Cost rating

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

Appendix B: Security classification of trial findings

OUTCOME: *Science GCSE (EAL pupils only)*

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

Threats to validity	Risk rating	Comments
Threat 1: Confounding	Moderate	Evidence of baseline imbalance on numerous variables with relevance to outcomes, but these do not appear to have impacted primary analysis based on present results.
Threat 2: Concurrent Interventions	Low	No evidence on differential uptake of concurrent interventions, and good discussion of usual practice.
Threat 3: Experimental effects	Low	No evidence of contamination.
Threat 4: Implementation fidelity	Moderate	Some aspects of low fidelity, especially around training. Limited data gathered on fidelity in classroom, and low sample sizes for surveys around this.
Threat 5: Missing Data	low	Very low levels of missing data, e.g. 3.5%.
Threat 6: Measurement of Outcomes	low	NPD outcomes, very robust, secure, low risk.
Threat 7: Selective reporting	low	All planned primary and secondary analyses reported, no evidence of selective reporting.

- **Initial padlock score:** 3 Padlocks. MDES (at randomisation) of .31 and high attrition (29%), so in accordance with EEF recommendation, choose lowest of three ratings = 3 padlocks.
- **Reason for adjustment for threats to validity**: [-0] Padlocks – Only two moderate threats, so no adjustment
- **Final padlock score:** initial score adjusted for threats to validity = 3 Padlocks

Appendix C: Effect size estimation

Appendix Table 2: Effect size estimation

Outcome	Unadjusted differences in means[1]	Adjusted differences in means	Intervention group n (missing)	Intervention group Variance of outcome[1]	Control group n (missing)	Control group Variance of outcome[1]	Pooled variance[2]	Population variance (if applicable)
Primary outcome, science, EAL pupils only	-0.003	0.10	532 (18)	3.29	504 (17)	3.67	2.46 (2.27, 2.67)	-NA-
Secondary outcome, history, EAL pupils only	-0.05	0.07	429 (33)	4.47	450 (40)	4.46	3.13 (2.84, 3.44)	-NA-
Secondary outcome, English, EAL pupils only	-0.00	0.09	892 (17)	2.94	864 (13)	2.81	1.92 (1.80, 2.05)	-NA-
Primary outcome, science, EAL+FSM pupils only	-0.17	0.12	222[SDC] (13[SDC])	2.82	190[SDC] (10[SDC])	4.33	2.55 (2.23, 2.87)	-NA-

Note. Further detail on the estimated variance components is provided in the notes to the corresponding tables in Appendix O.

[1] $M_{(intervention)} - M_{(control)}$; Variance estimate: Average of raw variances from bootstrap runs. [2] Variance estimate: Residual-level and school-level variance of the respective estimated mixed model incl. covariates, CI estimated via bootstrap. [SDC]Please note that the exact number of missing values for FSM needed to be redacted for SDC; the N in this table is reported with reference to the adjusted N = 435 of pupils with FSM.

Further appendices

Please see separate technical report with further appendices:

- Appendix D: ‘EAL in the Mainstream Classroom’ pilot report
- Appendix E: Memorandum of Understanding/information sheet
- Appendix F: Headteacher and teacher consent forms
- Appendix G: Parent consent form
- Appendix H: Addendum to Memorandum of Understanding
- Appendix I: Updated GDPR parent information sheet and consent
- Appendix J: Data sharing agreement
- Appendix K: DfE proficiency in English bands
- Appendix L: Science teacher surveys
- Appendix M: Observation schedule
- Appendix N: Imbalance at baseline analyses and distributions of outcomes
- Appendix O: Results of model estimations

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

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

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