

INTERVENTION	Integrating English
DEVELOPER	Enfield Council
EVALUATOR	Sheffield Hallam University
TRIAL REGISTRATION NUMBER	ISRCTN44415239
TRIAL STATISTICIAN	Sean Demack/Martin Culliney
TRIAL CHIEF INVESTIGATOR	Mike Coldwell/Martin Culliney
SAP AUTHOR	Martin Culliney/Sean Demack
SAP VERSION	1
SAP VERSION DATE	June 2017
EEF DATE OF APPROVAL	
DEVELOPER DATE OF APPROVAL	

Protocol and SAP changes

Amended protocol submitted to EEF 18/05/2017. This reflects agreed changes to the outcome measure, which is now an additional assessment using KS2 Writing Past Papers. The original protocol stated that the primary outcome measure as KS2 Writing but changes to the marking scheme from 2016 render this measure unsuitable for this trial. Also, KS2 Reading has been added as a secondary outcome as it was discussed in previous discussions between EEF, the developer and the evaluation team, but was not included in the protocol.

Table of contents

Protocol and SAP changes	1
Table of contents.....	1
Introduction.....	2
Study design	2
Randomisation	3
Calculation of sample size	3

Follow-up.....	4
Outcome measures.....	4
Primary outcome.....	4
Secondary outcomes.....	5
Analysis.....	5
Primary intention-to-treat (ITT) analysis.....	6
Interim analyses.....	6
Imbalance at baseline for analysed groups.....	7
Missing data.....	7
Non-compliance with intervention.....	7
Additional analyses.....	7
Subgroup analyses.....	9
Effect size calculation.....	9
Report tables.....	9

Introduction

The Integrating English intervention is based on *Language in Learning Across the Curriculum* (henceforth LiLAC), a training course developed by Australian linguistics experts, widely used in Australia and owned by the Government of South Australia. The LiLAC course enables non-specialist teachers to adopt a functional approach to linguistics and grammar, aiming to break down the process of teaching language to pupils. It treats communication in academic subjects as a matter of 'learning how to mean' using the semiotic resources available to that subject: that is to say students learn through language, thus language learning is central to learning in all subjects. Social semiotic language programmes such as Integrating English encourage English language and literature teachers to focus on the grammatical features of verbal art and everyday rhetoric; science teachers to focus on the grammatical features of classification, experimentation and reporting; history teachers to focus on the grammatical features of recounts and causation; maths teachers to focus on the grammatical features of problems, explanations and proofs; and so on.

The LiLAC course is delivered by external trainers. There are four one-day accredited modules attended by teachers in person, with readings and practical homework tasks between each session. Training takes place off school site, in five regional centres. Teachers visit their nearest regional centre for the training.

LiLAC training is an integral part of the Integrating English programme, which also provides ongoing support via FRONTER, an online platform. During the latter stages of training, schemes of work are developed by trainees, mentors and Integrating English advisers, based on learning from the LiLAC programme. Teaching based on these schemes of work then takes place.

Study design

- Participants: KS2 pupils (Y5 in 2016/7), at schools with at least 8 EAL pupils in Y5 at that time
- Design: Two-arm, school-level clustered randomised controlled trial (CRCT)
- 91 primary schools recruited in five geographical hub areas (London & Herts; Birmingham; North East; Cambridgeshire & Norfolk; Leicestershire & Peterborough)

- Y5 and Y6 teachers in intervention schools to receive LiLAC training, all pupils in relevant cohorts to be taught by these staff during Summer 2016/7 and throughout 2017/8.
- Baseline testing in November 2016, outcome testing in June 2018.

Randomisation

- Unit of randomisation: school. School level randomisation is more practical for recruitment and implementation. It also eliminates the risk of overspill, which is crucial for ensuring robustness.
- Method: minimisation (five school-level factors were used: Mean KS2 attainment, % EAL pupils, mean EAL fluency scale score, Number of Y5 classes, geographical hub)
- Allocation: 46 intervention schools, 45 control schools
- Randomisation took place after baseline testing in November 2016. Schools informed of allocation on November 30th 2016.

Calculation of sample size

A power analysis was undertaken based on the Bloom (2007) formula and supported by the Optimal Design Software. The findings are presented in Table 1. The research design is a clustered randomised controlled trial incorporating two levels (pupils clustered into schools) with randomisation at the school level.

Using the Bloom formula allows baseline covariates to be added at the cluster level, pupil level or both. Having a covariate at both school and pupil levels maximises the precision of the trial, leading to a smaller MDES estimate than approaches that use only one covariate or none. This is indicated in the formula, where ρ = School level intra-cluster correlation (estimated as 0.14); $R_c^2 = R^2$ for cluster level covariate (estimated as $0.6^2 \sim 0.36$; $1 - R_c^2 \sim 0.64$); $R_i^2 = R^2$ for pupil level covariate (estimated as $0.74^2 \sim 0.5476$ $1 - R_i^2 \sim 0.4524$); j = number of schools, estimated at between 60 & 100 in protocol, updated as 91 in SAP, 100 used in example below; m = number of participants per school, estimated as 40 per school for analyses involving all pupils and as 10 per school for EAL subsample analyses; $Mn-k*-2$ = z-score multiplier (=2.83 for $p < 0.05$ & power=0.80); P = proportion of schools assigned to intervention (=0.5; $1 - P = 0.5$).

Equation 1: Bloom et al (2007) MDES equation

$$MDES = M_{n-k*-2} \sqrt{\left(\frac{\rho(1 - R_c^2)}{P(1 - P)j} \right) + \left(\frac{(1 - \rho)(1 - R_i^2)}{P(1 - P)jm} \right)}$$

The power analysis presents the estimated Minimum Detectable Effect Sizes (MDES) for the primary outcome (writing assessment). For these MDES estimates, a statistical power of 0.8 is adopted, assuming that 100 schools will be recruited.

A statistical power of 0.8 means that there is an 80% chance of detecting an effect (or difference) between the intervention and control group samples if a true effect exists. The Minimum Detectable Effect Size (MDES) is the smallest effect size that the research design would be able to detect as being statistically significant with a statistical power of 0.8. For example, an MDES of 0.23 indicates that a difference in the outcome scores of the intervention and control group of 0.23 standard deviations (or greater) would be identified as statistically significant with a statistical power of 0.8.

From the EEF 'Master Test' spreadsheet, the (participant level) correlation between GL Progress in English (predecessor to PTE) and KS2 attainment is estimated to be 0.74 from the EEF¹. We estimate the (school level) correlation between PTE and KS2 attainment more conservatively at 0.60.

Clustering at the school level is estimated using a school level Intra Cluster Correlation Coefficient (ICC) of 0.14², based upon the KS2 ICC statistics provided by the EEF. This means that we estimate that 14% of the variation in the primary outcome will lie at the school level and the remaining 86% will be at the individual pupil level.

Table 1 shows MDES estimates for an outcome only analysis and the analysis that includes baseline covariates at both participant (GL PTE) and school (GL PTE mean score) levels and is based on the following estimates:

- **Number of schools** **100**
- **Number of pupils (40 per school)** **4,000**
- **School level ICC** **0.14**
- **Pupil-level correlation between PTE & primary outcome (r_i)** **0.74**
- **School-level correlation between PTE & primary outcome (r_s)** **0.60**

Table 1: Estimated Minimum Detectable Effect Size (MDES) for planned analyses for the primary outcome of the Integrating English clustered RCT

Analyses including ALL pupils (regardless of EAL status) - estimated as 40 pupils per school			
	100 schools	80 schools	60 schools
Outcome Only	0.23 standard deviations	0.26	0.30
Including baseline covariates ($r_i=0.74$; $r_s=0.60$)	0.18	0.20	0.23
Analyses of EAL pupils - estimated at 10 pupils per school			
	100 schools	80 schools	60 schools
Outcome Only	0.27 standard deviations	0.30	0.35
Including baseline covariates ($r_i=0.74$; $r_s=0.60$)	0.20	0.23	0.26

Updated with final trial school numbers

The final number of schools recruited to the trial was 91. Assuming the same details shown in Table 1, this results in MDES estimates of 0.19 (all pupils) and 0.21 (EAL pupils).

Follow-up

The outcomes for this trial will be collected at the end of the 2017/18 academic year so no details on missing data is available for this SAP.

Outcome measures

Primary outcome

The primary outcome is Writing, measured by raw scores on past KS2 Writing papers.

¹ Based on correlation between the GL PiE test (which the GL PTE test has since replaced) and the previous KS2. Taken from EEF 'Master Test Database', using NPD data 2013-2014.

²

https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol/ICC_2015.pdf

This assessment will be administered by the evaluators in June 2018. Invigilation and marking services will be provided by external contractors and will be blinded.

Secondary outcomes

The secondary outcomes are raw score on KS2 Grammar, Punctuation and Spelling (NPD field: KS2_GPSMARK, 2017/8) and KS2 Reading (NPD field: KS2_READMARK, 2017/8). The inclusion of both test scores will ensure that all aspects of participating pupils' English are covered in either the primary or follow on analyses.

Analysis

Table 2 below summarises the planned impact analyses for the *Integrating English* clustered RCT. The primary measure outcome is KS2 Writing, to be assessed through a test administered by the evaluators (KS2 past papers), the two secondary outcome measures are KS2 Grammar, Punctuation and Spelling (GPS), and KS2 Reading. The impact analyses will first include all pupils. It will then focus solely on EAL pupils. The five point fluency scale introduced in autumn 2016 (which was not available during the trial recruitment period) will be used in sensitivity analyses as detailed below.

A multilevel approach will be taken, with pupils clustered into schools. Multilevel linear regression models will be constructed for the primary and secondary outcome measures.

Recognising the EEF requirement to control for prior attainment and peer review comments, we plan to construct two models including prior attainment. The first prior-attainment model (Stage 1.1) will include the GL PTE baseline (school centred) as a covariate at the pupil level. We are undertaking a baseline test (GL PTE level 9) rather than using pupil level KS1 data given the higher level of missing KS1 data for EAL pupils³. The second prior-attainment model (Stage 1.2) will include both the pupil-level, school-centred GL PTE measure and school-level mean GL PTE scores as covariates. Including baseline covariates at both school and pupil levels results in lower MDES estimates than would be obtained with a single covariate at one of these levels. In other words, for the same level of statistical significance ($p < 0.05$) and statistical power (80%), a model that includes both school and pupil level covariates will be able to detect a smaller effect size as statistically significant than a model that just included one covariate at either school or pupil levels.

The model that includes prior attainment at just the pupil level is more reflective of approaches taken in other EEF evaluations and therefore the trial findings will be more directly comparable. The second stage that includes prior attainment at both pupil and school levels is included to reflect the original research design and power analyses for this trial. We will report both models in the final report.

Five school-level variables⁴ were used within the minimisation approach that was used to allocate the 91 schools into the Integrating English intervention ($n=46$) and control ($n=45$) groups. Minimisation was used to try to ensure a good baseline balance between the two groups - and Table 2 in the appendix shows that, at the school level, a good baseline balance across all five minimisation variables and two additional variables (KS1 attainment and %FSM). Balance at the pupil level will be explored when this data is available.

The five school-level minimisation variables will be included into the model at stage 2. The purpose of this stage is to acknowledge the use of these variables in minimisation and to examine if and how their inclusion changes the estimated effect size for the impact of Integrating English. Including minimisation variables should increase the precision of the estimates, potentially reducing the confidence intervals. As Model 1.2 follows more closely the design decisions included in the power

3

https://v1.educationendowmentfoundation.org.uk/uploads/pdf/EAL_and_educational_achievement2.pdf

⁴ KS2 attainment, %EAL, Number of Y5 classes, geographical hub area & EAL language proficiency.

calculation, the Model 2 will add the minimisation variables to Model 1.2 only. Model 2 will be considered the headline result.

The secondary outcome measures are KS2 Grammar, Punctuation and Spelling (GPS), and KS2 Reading. The impact analyses for this secondary outcome will follow exactly the same multilevel modelling steps shown in Table 2 for the primary outcome.

Example of STATA SYNTAX for Integrating English impact analyses

*** Stage 1.1 (Simplest model)**

```
.mixed OUTCOME Group dummy Pupil-level GL PTE 9 Score (centred) || School level Identifier:
estat icc
```

*** Stage 1.2**

```
.mixed OUTCOME Group dummy Pupil-level GL PTE 9 Score (centred) School-level mean GL
PTE 9 Score || School level Identifier:
estat icc
```

*** Stage 2 (Precise Model)**

```
.mixed OUTCOME Group dummy Pupil-level GL PTE 9 Score (centred) School-level mean GL
PTE 9 Score [school level variables used for minimisation] || School level Identifier:
estat icc
```

Primary intention-to-treat (ITT) analysis

Table 2: Summary of primary ITT analysis plan. Multilevel linear regression with two levels (pupils clustered into schools). This will be used for analysis of both the primary and secondary outcomes.

Impact Analysis for Primary and Secondary Outcomes			
Sample	Approach	Level 1 (pupil) covariates	Level 2 (school) covariates
Complete Sample	Intention to treat	GL PTE level 9 score, school centred	Dummy (1=intervention; 0=control)
		GL PTE level 9 score, school centred	Dummy (1=intervention; 0=control), Mean GL PTE 9 Score
		GL PTE level 9 score, school centred	Dummy (1=intervention; 0=control), Mean GL PTE 9 Score, % EAL pupils, mean EAL fluency scale score, Number of Y5 classes, geographical hub
EAL Pupils only ²	Intention to treat	GL PTE level 9 score, school centred	Dummy (1=intervention; 0=control)
		GL PTE level 9 score, school centred	Dummy (1=intervention; 0=control), Mean GL PTE 9 Score
		GL PTE level 9 score, school centred	Dummy (1=intervention; 0=control), Mean GL PTE 9 Score, % EAL pupils, mean EAL fluency scale score, Number of Y5 classes, geographical hub

Notes on analysis plan:

¹As only the binary EAL measure was available during the trial recruitment period, this will be used for the subgroup analysis.

²Within the exploratory analyses we will use the new five point EAL language proficiency measure introduced in 2016/17.

Interim analyses

Not applicable to this trial.

Imbalance at baseline for analysed groups

The intervention and control groups will be compared using both school and pupil-level baseline details including GL PTE Level 9 score (at pupil level and school level mean), FSM (The NPD field 'EVERFSM_6_p_' is to be used as the indicator of FSM) and EAL status. These analyses will provide detail on whether the school level randomisation resulted in a good balance at the pupil level between the intervention and control groups.

Missing data

There are no plans to replace missing data using imputation. Instead, we will examine missing data for pattern and explore the potential impact of missing data on our estimated effect sizes.

The pattern of missing data will be analysed with respect to a range of pupil-level variables (such as the trial outcomes, PTE Level 9 score, gender, FSM and EAL status). For each predictor, we will compare missing cases with non-missing cases on the outcome measure. We will also run a logistic regression model with non/missing outcome data as the model outcome and the pupil-level covariates listed in Table 2 as predictors. These analyses will help to detect patterns in the missing data.

To examine the extent to which any missing data could alter the estimated impact of the Integrating English intervention, we will draw on the work of Gorard (Gorard 2015; Gorard and Gorard 2016⁵) to examine the number of cases needed to disturb a finding. Specifically, the number of cases needed in order to reduce an observed effect size to zero will be calculated for the primary and secondary outcomes.

Our planned analyses of missing data will be used enable a more comprehensive interpretation on the security of the effect size estimates within the primary and secondary analyses.

Non-compliance with intervention

The four fidelity measures to be used are:

- Attendance at training events; measured by number of sessions
- Attainment of certificate for completing the coursework; binary measure
- Scheme of Work; ranked from 0 (none) to 3 (excellent) by process evaluation team
- Project Action Plan; ranked from 0 (none) to 3 (excellent)

Discussions as to how these will be combined into a single measure of fidelity are still ongoing. The SAP and protocol will be updated once this is finalised.

The statistical association between fidelity (attendance of training, completion of homework, preparation of schemes of work) and the trial outcomes will be examined. If the association is found to be statistically significant, an exploratory CACE analysis will be undertaken - as illustrated in Table 4 below.

Additional analyses

As specified in the protocol, further exploratory analyses will ascertain the impact of the *Integrating English* programme on pupils classified as FSM⁶ through subgroup analysis and will also explore the

⁵ Gorard, S. (2015) 'Rethinking `quantitative' methods and the development of new researchers', *Review of education*, 3 (1). pp. 72-96

Gorard, S. and Gorard, J. (2016) 'What to do instead of significance testing? Calculating the `number of counterfactual cases needed to disturb a finding' *International journal of social research methodology*. 19 (4). pp. 481-490

⁶ FSM is a measure to identify pupils claiming Free School Meals. The NPD 'FSMEver_6_p' variable, indicating pupils who have ever been eligible for FSM at any time, will be used for these analyses.

relationship between EAL status and the two outcome measures for both the complete sample and the FSM subgroup. These analyses will follow the specification used for the headline figure (Model 2) and will include an additional five-point EAL language fluency measure as a pupil level covariate.

Table 3: Summary of exploratory analysis plan. Multilevel linear regression with two levels (pupils clustered into schools). This will be used for analysis of both the primary and secondary outcomes.

<i>Exploratory ITT Analyses for Primary and Secondary Outcomes</i>			
Sample	Approach	Level 1 (pupil) covariates	Level 2 (school) covariates
Complete Sample	Intention to treat	GL PTE level 9 score, school centred EAL fluency status	Dummy (1=intervention; 0=control)
		GL PTE level 9 score, school centred EAL fluency status	Dummy (1=intervention; 0=control), Mean GL PTE 9 Score
		GL PTE level 9 score, school centred EAL fluency status	Dummy (1=intervention; 0=control), Mean GL PTE 9 Score, % EAL pupils, mean EAL fluency scale score, Number of Y5 classes, geographical hub
FSM pupils only	Intention to treat	GL PTE level 9 score, school centred	Dummy (1=intervention; 0=control)
		GL PTE level 9 score, school centred	Dummy (1=intervention; 0=control), Mean GL PTE 9 Score
		GL PTE level 9 score, school centred	Dummy (1=intervention; 0=control), Mean GL PTE 9 Score, % EAL pupils, mean EAL fluency scale score, Number of Y5 classes, geographical hub
		GL PTE level 9 score, school centred EAL fluency status	Dummy (1=intervention; 0=control)
		GL PTE level 9 score, school centred EAL fluency status	Dummy (1=intervention; 0=control), Mean GL PTE 9 Score
		GL PTE level 9 score, school centred EAL fluency status	Dummy (1=intervention; 0=control), Mean GL PTE 9 Score, % EAL pupils, mean EAL fluency scale score, Number of Y5 classes, geographical hub

Notes on analysis plan:

¹As only the binary EAL measure was available during the trial recruitment period, this will be used for the subgroup analysis.

²Within the exploratory analyses we will use the new five point EAL language proficiency measure introduced in 2016/17.

Table 4: Summary of CACE analysis plan

<i>CACE Analyses for Primary and Secondary Outcomes</i>			
Sample	Approach	Level 1 (pupil) covariates	Level 2 (school) covariates
Complete Sample	CACE	GL PTE level 9 score, school centred	Dummy (1=intervention; 0=control)
		GL PTE level 9 score, school centred	Dummy (1=intervention; 0=control), Mean GL PTE 9 Score
		GL PTE level 9 score, school centred	Dummy (1=intervention; 0=control), Mean GL PTE 9 Score, % EAL pupils, mean EAL fluency scale score, Number of Y5 classes, geographical hub
EAL Pupils only	CACE	GL PTE level 9 score, school centred	Dummy (1=intervention; 0=control)
		GL PTE level 9 score, school centred	Dummy (1=intervention; 0=control), Mean GL PTE 9 Score
		GL PTE level 9 score, school centred	Dummy (1=intervention; 0=control), Mean GL PTE 9 Score, % EAL pupils, mean EAL fluency scale score, Number of Y5 classes, geographical hub

Subgroup analyses

Subgroup analyses specified in the protocol - detailed in table 3 (above). This will be conducted using separate datasets containing only FSM or EAL pupils.

Effect size calculation

The impact of Integrating English will be measured using the Hedges g effect size statistic based on the formula shown below.

Equation 2: Hedges g effect size calculation

$$ES = \frac{(T - C)_{adjusted}}{\sqrt{\delta_s^2 + \delta_p^2}}$$

Where δ_s^2 is the school level variance and δ_p^2 is the pupil level variance and $(T - C)_{adjusted}$ is the coefficient estimate for the group identifier dummy variable from the multilevel model which is adjusted by other covariates included in each stage.

The upper and lower confidence intervals for the coefficient estimate for the group identifier dummy variable will also be standardised into Hedges g effect size statistics.

The headline effect size will be that obtained in the regression model labelled above as 'Stage 2'. Effect sizes and confidence intervals will be calculated for all model stages.

Report tables

Table 1: Post-Minimisation School level comparison of Intervention and control groups*
Number of Y5 classes, language proficiency levels & geographical hub area.

Variable	Intervention	Control
Number of Y5 classes (data gathered from schools)		
Mean (sd)	2.04 (0.59)	2.09 (0.93)
Min Max	1.0: 3.0	1.0 : 5.0
EAL Language Proficiency (data gathered from schools)		
Proficiency A (New to English; numbers of pupils per school scored as 'A')		
Mean (sd)	1.6 (2.69)	1.5 (3.35)
Min Max	0 : 15	0: 17
Proficiency B (Early Acquisition)		
Mean (sd)	3.5 (6.43)	2.8 (4.88)
Min Max	0 : 33	0 : 30
Proficiency C - (Developing Competence)		
Mean (sd)	7.9 (9.92)	7.6 (8.30)
Min Max	0 : 48	0 : 39
Proficiency DE (Competent / Fluent)		
Mean (sd)	16.8 (15.15)	16.8 (18.00)
Min Max	0 : 70	0 : 95
Proficiency Details Missing		
Mean (sd)	0.3 (0.75)	0.4 (1.12)
Min Max	0 : 4	0 : 5
Not EAL		
Mean (sd)	26.2 (17.33)	23.6 (17.07)
Min Max	0 : 70	0 : 83
Geographical Hub Area		
London & Herts	13 (28%)	10 (22%)
Birmingham	11 (24%)	11 (24%)
North East	8 (17%)	7 (16%)
Cambs & Norfolk	7 (15%)	11 (24%)
Leics-Peterboro	7 (15%)	6 (13%)
KS2 Attainment (TAPS - 2015/16 KS2 school census)		
Mean (sd)	28.2 (1.47)	28.4 (1.58)
Min Max	24.8: 31.5	24.7 : 31.4
n=	46	45
%EAL (PTEALGRP2- 2015/16 KS2 school census)		
Mean (sd)	45.6% (24.06 pps)	45.7% (22.64 pps)
Min Max	9% to 91%	11% : 98%
n=	46	45
%FSM (%FSM in last 6 years; PTFSM6CLA1A - 2015/16 KS2 school census)		
Mean (sd)	43.8% (17.13 pps)	43.6% (19.06 pps)
Min Max	7% : 75%	16% : 87%
n=	46	45
KS1 attainment for KS2 cohort (TKS1APS - 2015/16 KS2 school census)		
Mean (sd)	14.7 (1.32)	14.6 (1.31)
Min Max	11.7 : 17.4	11.4 : 16.7
n=	46	45

* data gathered from schools prior to randomisation or obtained from the 2015/16 KS2 school census.