Flexible Phonics
Evaluation Protocol

Evaluator (institution): Institute for Employment Studies
Principal investigator(s): Anneka Dawson and Helen Gray

Education Endowment Foundation

PROJECT TITLE	Flexible phonics, a two-armed cluster randomised trial
DEVELOPER (INSTITUTION)	UCL, IOE
EVALUATOR (INSTITUTION)	Institute for Employment Studies (IES)
PRINCIPAL INVESTIGATOR(S)	Anneka Dawson and Helen Gray
PROTOCOL AUTHOR(S)	Anneka Dawson, Helen Gray and Clare Huxley
TRIAL DESIGN	Two-arm cluster randomised controlled trial with random allocation at school level
TRIAL TYPE	Efficacy
PUPIL AGE RANGE AND KEY STAGE	Reception year age 4-5
NUMBER OF SCHOOLS	100
NUMBER OF PUPILS	2,300
PRIMARY OUTCOME MEASURE AND SOURCE	York Assessment for Reading Comprehension Test- word recognition subscale
SECONDARY OUTCOME MEASURE AND SOURCE	York Assessment for Reading Comprehension Test full score, Mispronunciation Communication Test, delayed post- test of Phonics screening test at end of Year 1

Protocol version history

VERSION	DATE	REASON FOR REVISION
1.2 [*latest*]		
1.1		
1.0 [*original*]		N/A

Table of contents

Study rationale and background

Studies show that systematic phonics is effective in supporting younger readers to master the basics of reading[1]. The Flexible Phonics approach is an optimisation of phonics by linking phonics to lexical and semantic information (direct mapping) and strategies to allow independence in reading of the 'deep' (irregular) orthography of English which admits many exceptions to phonic rules (set-for-variability). This approach aims to reinforce phonics learning through reading words in a meaningful context such as reading children's books as well as teaching children strategies to help with learning new, irregular words.

Evidence suggests that combining phonics teaching and book reading is more effective than teaching phonics alone. A series of studies by Hatcher and colleagues (Hatcher et al., 1994, 2004, 2006) found that interventions where children were taught phonics and then read authentic texts in the same session were more effective than teaching phonics alone.

Further studies have found that explicitly linking phonics learning with a relevant reading task was more effective than regular phonics teaching or a vocabulary learning task. Shapiro and Solity (2008) taught phonics to children aged 6-7 years and then explicitly linked this to reading selected children's books which contained a high density of grapheme-phoneme mappings that had been taught as part of the phonics. This approach improved reading outcomes over regular phonics teaching. Similarly, Chen and Savage (2014) established in an RCT with children aged 7-8 years that teaching an explicit direct mapping approach was more effective than a control condition where children were taught a vocabulary intervention. In the direct mapping condition, children articulated grapheme-phoneme mapping that they had recently learned as part of the shared reading of children's books.

The teaching of phonic strategies has been shown to positively affect reading outcomes (Savage et al., 2007) and several studies have shown that focusing on variable vowel pronunciation positively impacts learning (Lovett et al., 2014; Savage & Stuart, 2001, 2006). Tunmer and Chapman (2012) demonstrated that phonics decoding skills can be broken down into component subskills. In current best practice synthetic phonics programme, children are taught to blend speech sounds to read full words, e.g. 'c'-'a'-'t' to read 'cat' and 'c'-'a'-'tch' to read 'catch'. However, some models of word-reading propose an additional step after blending where learners compare the blended sounds with words known to them in their mental lexicon (i.e. existing vocabulary). In particular, additional processing applied in cases where there is variation in the pronunciation of vowels, i.e. irregular words, has been identified in models as either 'set for diversity' (Gibson, 1965) or 'set for variability' (Venezky, 1999).

Recent studies have found that phonics approaches that explicitly teach strategies for set-for-variability are more effective than standard phonics testing. A recent experimental study (Steacy et al., 2016) found that phonics teaching that included a focus on variable pronunciations of vowels (i.e. set-for-variability) yielded better reading outcomes (specifically on the pronunciation of words with variable vowels) compared with phonics teaching that did not incorporate this aspect. Several studies have found that teaching set-for-variability as a strategy for correcting irregular words that have been incorrectly pronounced with a regularised pronunciation improved children's ability to self-correct when they attempt to read new irregular words (Dyson at al., 2017; Zipke, 2016). Furthermore, several studies have proposed that this additional processing step can be applied to all words including words with regular pronunciation (Elbro et al.,2012; Elbro & De Jong, 2017; Kearns at al.,

[1] E.g. Torgerson et al., 2018; Camilli et al., 2008; Galuschka et al., 2014

2016) which suggests that teaching set-for-variability strategies may help with reading of all words and that there may be longitudinal effects seen on development (Steacy et al, 2019).

A Quasi-Experimental Design (QED) study in Canada found that an intervention combining these two strategies (direct mapping and set-for-variability) was more effective than best practice phonics teaching when taught to struggling readers aged 5-7 years, with additional positive outcomes for spelling and reading (Savage et al., 2018). As phonics teaching starts at an earlier age in England, the author suggests that the children in the Canadian study would be at a similar developmental phase of reading as children aged 4-5 years in England. The current study will use a randomised controlled trial (RCT) to test whether a similar intervention incorporating direct mapping and set-for-variability approaches would lead to improved reading outcomes for children of all ability compared with current best practice phonics teaching in England. Further to this, as the intervention in Canada was taught by research assistants, the current study will also investigate the feasibility of reception teachers and teaching assistants delivering this intervention as part of everyday teaching. As phonics teaching in Reception is mixed within schools (and can be led by teachers or teaching assistants in different sized groups) a cluster RCT is the most appropriate design so that contamination across teachers/ groups does not take place. The context of phonics teaching in England is summarised below.

Systematic synthetic phonics has been a core component of teaching reading in England for the last decade since the 2010 white paper 'The Importance of Teaching' was published by the then Conservative and Liberal Democrat Government. In the report, the Government recommended systematic synthetic phonics as the 'best method' for teaching reading, and pledged to promote this approach through providing resources to promote the teaching of systematic synthetic phonics in all primary schools and by making it a part of teacher training. The paper also recommended assessing children's reading at age six through a test of pupil's ability to decode words, and the phonics screening check for children in Year 1 was piloted in 2011 and rolled out across England in 2012 (DfE, 2011). In 2013, the Government published guidance for eight phonics programmes which set-out how the programmes met DfE's criteria for effective systematic synthetics teaching programmes (DfE, 2014). These programmes have now become well-established among primary schools in England and children are expected to demonstrate phonic knowledge as part of the Early Years Framework (DfE, 2017). The Flexible Phonics intervention aims to build on current best practice by training Reception teachers to apply new approaches within phonics teaching (Direct Mapping and Set-for-Variability) which can help children with reading new irregular words

As well as potential benefits to children's reading and to current phonics practice in the UK, this study makes an especially valuable contribution to the evidence base that EEF is developing. At the time of commissioning this evaluation, the EEF had funded 10 phonics projects but none had focused specifically on Reception class learners. The Flexible Phonics evaluation would therefore fill a gap in EEF's phonics portfolio. Further to this, the Flexible Phonics study contributes to a stated priority of the Early Years Professional Development round which was to improve the training of Reception teachers.

Intervention

The Flexible Phonics intervention helps Reception class teachers and Teaching Assistants in the classroom delivery of new strategies designed to optimise the teaching of reading to all children. The work fits well around existing phonics programmes that can be delivered broadly as usual. A novel aspect of *Flexible Phonics* is that it gives children more strategies

to flexibly read all words, and could be particularly powerful in enabling children to independently read novel exception words (words that break phonic rules such as 'the', 'two', 'between', 'above', etc.). Children learn how to use phonics in close conjunction with authentic children's texts to become confident, motivated, readers. The logic model for the intervention is set out on page 7.

The TIDieR framework for the intervention is as follows:
Name: Flexible Phonics:
Why: Systemic phonics now has a lot of evidence but there are still ways to refine it further and recent evidence suggests combining direct mapping and set-for-variability strategies will help to do this.
Who (recipients): All pupils in Reception year (age 4-5) but there may be added benefits for low achieving pupils. All Reception class Teachers and Teaching Assistants (TAs) will be the direct recipients of the training and then will deliver to the pupils in lesson time.
What (materials): Those who are allocated to receive the intervention will receive three half days of professional development training (remote training using a virtual meeting platform such as Zoom), a copy of a Teacher Manual and associated resources (which are still under development in an initial pilot phase) and two in-class follow up visits where research assistants (with teacher's consent) will observe the classroom context and provide further feedback and guidance around delivering the intervention (these may also need to be virtual depending on the status of the covid-19 pandemic). They will also receive free children's books to the value of £400 per school which can be used to implement the strategies. Ongoing phone and email support will be provided if needed.
What (procedures): The training will introduce the two strategies for the Teachers and TAs to implement in their teaching as follows:

1. The first strategy, Direct Mapping, requires children to read texts that include several examples of the GPCs that they have just learned. In the first instance, these will be carefully selected pre-existing decodable texts, or specifically crafted controlled texts before real books are introduced slowly and strategically. While many models of phonics teaching link phonics and texts, DM aims to do so more thoroughly, consistently, and on the same day as children learn the specific GPCs, aiming to ensure that children understand phonics in context

2. The second strategy, Set-for-Variability (SfV), explicitly teaches pupils to add in another step after they have blended phonemes to graphemes where pupils 'set-for-variability'. This is a metacognitive step, where pupils consider what the word may be, given both the distance between these blended sounds and known words, and potential spelling to sound inconsistencies. For example, when they sound out the phonemes 'c'-'a'-'t', the sounds they make bear little resemblance to the actual word 'cat'. SfV encourages pupils to take a moment to consider what the word may be from the words that they know. This enables children to better recognise all words but can also be especially useful when learning to recognise exception words (e.g. 'wasp'). In comparison with other phonics programmes, SfV makes this metacognitive step following blending much more explicit and can enable children to be more flexible when approaching difficult words

Who (provider): Professor Savage and his team at UCL IOE will deliver the training and follow- up sessions to the Teachers and TAs who will in turn deliver the strategies within their normal phonics practice (both in whole class and small group delivery) after children have learned grapheme to phoneme correspondences (GPCs). (A phoneme is the smallest unit of sound e.g. the word 'rain' has 3 phonemes; 'r'-'ai'-'n'. A grapheme is the written symbol that represents that sound, which can be a single letter or a sequence of letters).
How (format): the strategies will be delivered in normal phonics lessons.
Where (location): The schools will be recruited from greater London

When and how much (dosage): The intervention is being delivered from January 2021- end of May 2021. The expectation is that all phonics lessons will incorporate the strategies during this time which is normally 3-4 times a week depending on the school.
Adaptation: Teachers will tailor and differentiate the content to suit children. There is freedom for teachers to adapt and modify as they go (although there will be a defined core that they must follow).
Control condition: The control condition is business as usual phonics practice and schools allocated to the control condition will receive £1000 at the end of the academic year when post-testing is complete.

Flexible Phonics
Evaluation Protocol

Evaluator (institution): Institute for Employment Studies
Principal investigator(s): Anneka Dawson and Helen Gray

Education Endowment Foundation

RATIONALE / NEED FOR INTERVENTION

Systemic phonics now has a lot of evidence but there are still ways to refine it further and recent evidence suggests combining direct mapping and set-for- variability strategies will help to do this.
1. The first strategy, Direct Mapping, requires children to read texts that include several examples of the GPCs that they have just learned on the same day. For children who cannot decode, these will be carefully selected pre-existing decodable texts, or if children can decode then specifically crafted controlled texts before real books are introduced slowly and strategically.
2. The second strategy, Set-for-Variability, explicitly teaches pupils to add in another step after they have blended phonemes to graphemes where pupils 'set-for-variability'. This is a metacognitive step, where pupils consider what the word may be, given both the distance between these blended sounds and known words, and potential spelling to sound inconsistencies. This enables children to better recognise all words but can also be especially useful when learning to recognise exception words

Long term outputs/ Impacts

- Improved overall literacy outcomes for pupils
- Improved phonological awareness for pupils
- Increased use of direct mapping and set- for variability strategies by pupils

Theory of change

To provide two strategies that Teachers and TAs can use in all Reception phonics teaching by supporting them to make careful modifications of their existing lesson plans to incorporate key Flexible Phonics strategies and approaches to add value to their standard practice which can improve children's literacy outcomes.

Inputs

- Selection of 115 schools from the Greater London.
- Pupil details collected (by evaluator for trial only).
- Time needed to attend three interactive half days of online training using a virtual meeting platform and then for research assistants to attend two follow-up sessions in school.
- Time needed in standard phonics lessons for new strategies
- £400 of specifically chosen books are given to the schools to enable the direct- mapping strategy

Activities

- All teacher and TAs attend three half days of interactive online training using a virtual meeting platform such as Zoom
- Teachers and TAs deliver Flexible Phonics strategies across all normal phonics lessons in January- May 2021
- In two follow- up sessions, research assistants observe a phonics lesson and afterwards provide support
- Teachers and TAs receive a manual and additional resources (developed during the pilot phase) to support the lessons
- Ongoing support is provided by

Outputs

- Teachers and TAs deliver the Flexible phonics strategies approx. 3-4 times a week as part of their standard practice
- Pupils use these strategies for everyday reading in school and at home

Short term outcomes/ Mediators

- Teachers/ TAs report more confidence in delivering phonics lessons
- Pupils have greater word reading abilities

Enabling factors / conditions for success

- Flexible Phonics strategies may be more or less effective with pupils from low income families (evaluated by free schools meals sub-group analysis)
- Flexible Phonics strategies may be more or less effective with pupils with higher or lower existing reading ability (evaluated by examination of low and high baseline assessment sub- group analysis)
- Spillover- effect should not be a problem for this trial as it is cluster randomised but can be confirmed through the IPE
- Geographical region may also have an impact on intervention effectiveness due to different pupil demographics, but this will not be examined in the current trial

Impact evaluation

Research questions

The primary research question is:

- RQ1. Does the Flexible Phonics intervention improve Reception children's word reading ability? (measured by the York Assessment for Reading Comprehension (YARC) Early Word Recognition subscale)

The secondary research questions are:

- RQ2. Does the Flexible Phonics intervention improve Reception children's literacy outcomes? (measured by more general literacy tests)
- RQ3. What is the differential impact of direct mapping and set- for-variability skills on children's word reading ability?
- RQ4. Does the Flexible Phonics intervention provide value- added improvement to Reception children's word reading ability compared to good phonics teaching alone in schools identified with good phonics practice?
- RQ5. Does the Flexible Phonics intervention improve word reading ability differentially for children eligible for Free School Meals (FSM)?
- RQ6. Does the Flexible Phonics intervention improve word reading ability differentially for children of low ability?
- RQ7. Does the Flexible Phonics intervention improve Reception children's phonics skills one year later at the end of Year 1?

Design

Table 1: Trial design

Trial design, including number of arms		Two-arm, cluster randomised control efficacy trial with pupil-level outcomes
Unit of randomisation		School
Stratification variables (if applicable)		None
Primary outcome	variable	Early Word Recognition
	measure (instrument, scale, source)	Early Word Recognition subscale raw score (0-30) from the York Assessment for Reading Comprehension (YARC)
Secondary outcome(s)	variable(s)	Early Word Reading composite measure Mispronunciation Correction Literacy over the longer-term
	measure(s) (instrument, scale, source)	For literacy: - The sum of standardised scores derived from each of the four YARC subscales i.e. early word recognition, letter sound knowledge, sound deletion and sound isolation. - Score on the Year 1 Phonics Screening check for longer-term outcomes. For Mispronunciation Correction: - An adapted version of Tunmer and Chapman's Mispronunciation Correction Test (2012) as used in Dyson et al. (2017) using the words most commonly used in English children's books.
Baseline for primary outcome	**variable**	Early Word Recognition
	measure (instrument, scale, source)	Early Word Recognition subscale raw score from YARC
Baseline for secondary outcome	**Variable**	Early Word Recognition and Letter Sound Knowledge composite measure
	measure (instrument, scale, source)	Constructed from the standardised scores for the Early Word Recognition and Letter Sound Knowledge subscales from YARC

As the intervention involves training all reception teachers and teaching assistants, randomisation will be at the school-level. Half of the participating schools will be allocated to treatment, the other half will be assigned to the control group. These latter schools will continue usual approaches to phonics teaching[2].

The study will measure the impact of flexible phonics for pupils in Reception class at the start of 2020/21. Children will participate in a pre-test of reading ability (the York Assessment

[2] While it is possible that not all Reception teachers and TAs will be trained in larger schools, within-school randomisation risks cross-contamination and demoralisation among teachers and teaching assistants randomised to the control group.

for Reading Comprehension, YARC using the early word recognition and letter sound knowledge subscales) prior to randomisation to verify treatment and control groups are well-matched. Information on the balance between the two groups pre-intervention enables assessment of the likely robustness of findings and accounts for some of the variance in the post-test, meaning sample size is optimised. In our TipsByText (TBT) trial we explored the possibility of avoiding a pre-test by gathering internal EYFS profile data from teachers, but found this infeasible. We also considered using the new Reception Baseline Assessment, but this has now been cancelled for autumn 2020 due to the covid-19 pandemic and so we are proceeding with using the YARC. We will manage the risk of delays due to the pre-test carefully (see risks section).

To reduce the costs of testing, as well as minimising the burden on schools, the pre-test and post-test will be administered to one class per school. Where there is more than one reception class per school, the class will be selected at random from a list of teachers provided by the school. Whilst only a single class will take part in the pre- and post-tests, as all teachers or TAs will be invited to take part in the training and the training will be designed to be cascaded to any non-participants, the transfer of teachers between classes should not affect whether pupils receive the intervention. However, it is possible there will be a dilution of effects which will be explored in the implementation and process evaluation through observations and interviews.

Randomisation

The aim is for 100 schools to participate in the trial. All recruited schools will be asked to supply a list of reception teachers. Where the school has more than one teacher of reception-aged children, one teacher per school will be selected at random from this list. Having made this initial selection of 100 teachers, each teacher will be assigned to the treatment or control group at random. The classes selected for the treatment and control conditions at the start of the 2020/21 academic year will be the focus in pre- and post-testing. Test administrators will not be told whether the school has been assigned to the treatment or control group.

Stata will be used to generate a unique random number for each reception teacher within each school with more than one reception class. Teachers will be sorted in ascending order on this random number and the teacher with the lowest random number from each school will be selected for random assignment.

Having made this initial selection of a single reception teacher from each school and carried out the pre-test, Stata will be used to generate a random number for each of the 100 teachers. Again, they will be sorted in ascending order on the random number and this will be used to derive a sort order variable. Those with odd numbers on the sort order variable will be assigned to the treatment group, and those with even numbers will be assigned to the control group. The delivery team will be supplied with a list of schools assigned to the treatment group in order to invite them to training.

The seed used to generate the random numbers used to select one teacher per school and to assign schools to the treatment and control groups will be saved, along with all other syntax used to make the random assignment. Analysts will not be blind to whether schools are part of the treatment or control group and which class is the subject of testing.

Participants

The intervention is targeted at children in Reception classes, who are expected to turn five in the 2020/2021 academic year. All children in Reception classes in the schools recruited to the trial will be eligible to participate. Parents will have the option of withdrawing their child's data from use in the trial. All teachers and teaching assistants of Reception-aged children in the schools assigned to the treatment group will be invited to attend training and participate in other activities to equip them to teach flexible phonics.

Schools participating in the English Hubs programme will not be eligible to participate in the trial. All other schools with Reception age children in Greater London will be eligible to participate as long as they are not participating in another EEF Reception year trial. All teachers and teaching assistants of Reception-age children allocated to the treatment group will be invited to participate in Flexible Phonics training and resources will be made available to those who are unable to attend.

The delivery team (from University College London, Institute of Education) are expected to recruit approx.115- 125 schools so that approx. 100 will go forward to participate in the trial allowing for attrition between signing up and being randomised.

Sample size calculations

Table 2: Sample size calculations

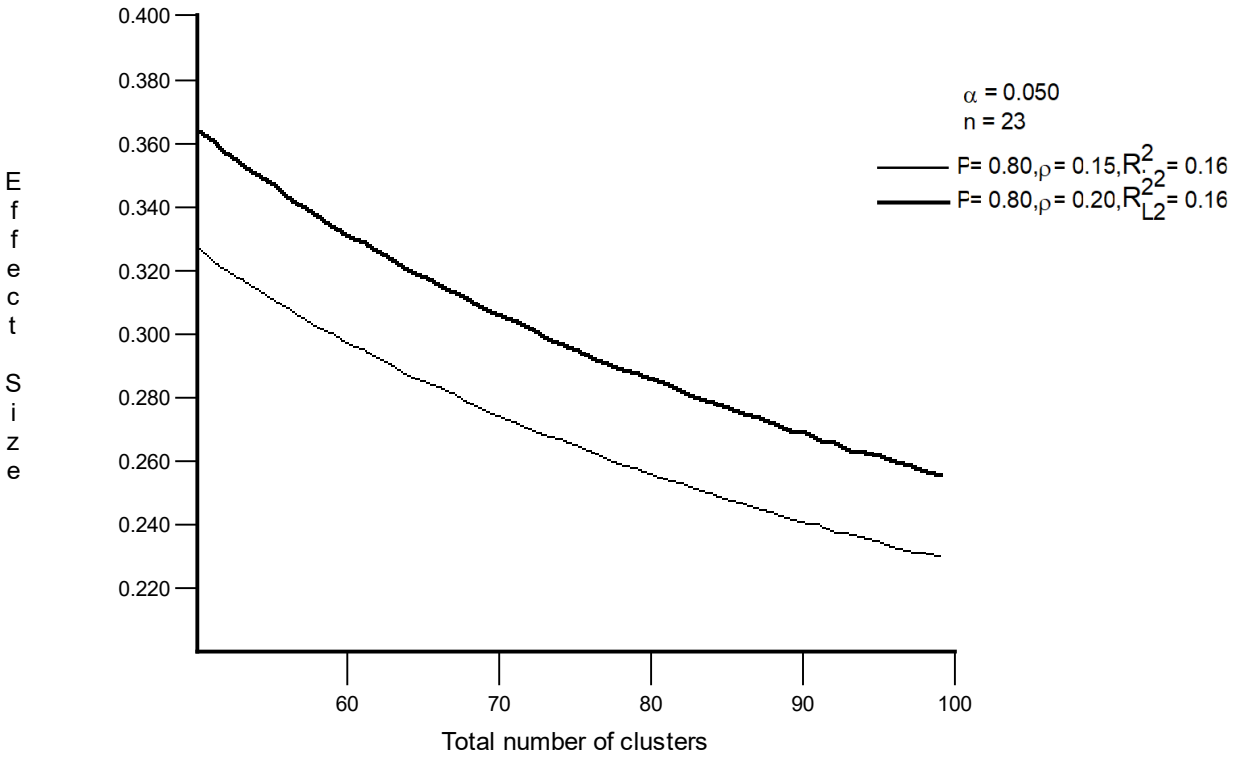
		OVERALL	FSM
Minimum Detectable Effect Size (MDES)		0.23 Standard deviations	.37 Standard deviations
Pre-test/ post-test correlations	level 1 (pupil)	0.4	0.4
	level 2 (class)		
	level 3 (school)		
Intracluster correlations (ICCs)	level 2 (class)		
	level 3 (school)	0.15	0.15
Alpha		0.05	0.05
Power		0.8	0.8
One-sided or two-sided?		Two-sided	Two-sided
Average cluster size		23[3]	3
Number of schools	Intervention	50	50
	Control	50	50
	Total	100	100
Number of pupils	Intervention	1,150	150
	Control	1,150	150
	Total	2,300	300

[3] This is based on the expected achieved sample size i.e. after attrition.

The MDE sizes reported in Table 2 were calculated using the Optimal Design software package, available at: http://hlmsoft.net/od/

DfE figures indicate that the average infant class contained 27 pupils in 2018[4]. If as many as 15 per cent of pupils withdraw from trial participation prior to randomisation or do not participate in data collection for the study's lifetime, this suggests an average of around 23 pupils per school will participate in the trial. The calculation of the MDE assumes that the pre-test explains 16 per cent of the variation in the post-test scores[5], and that the intra-class correlation is 0.15[6].

With the standard assumptions of 80% power and 5% significance level and an even split in the number of schools in the treatment and control groups between the 100 schools expected to participate in the trial, the MDE on the primary outcome measure would be 0.23 standard deviations. This calculation is based on an assumption that only pupils in a single class per school will participate in the pre- and post-tests. If the intra-class correlation is 0.20, rather than 0.15, but all the other assumptions stay the same, the MDE increases to 0.26 standard deviations (see graph below).

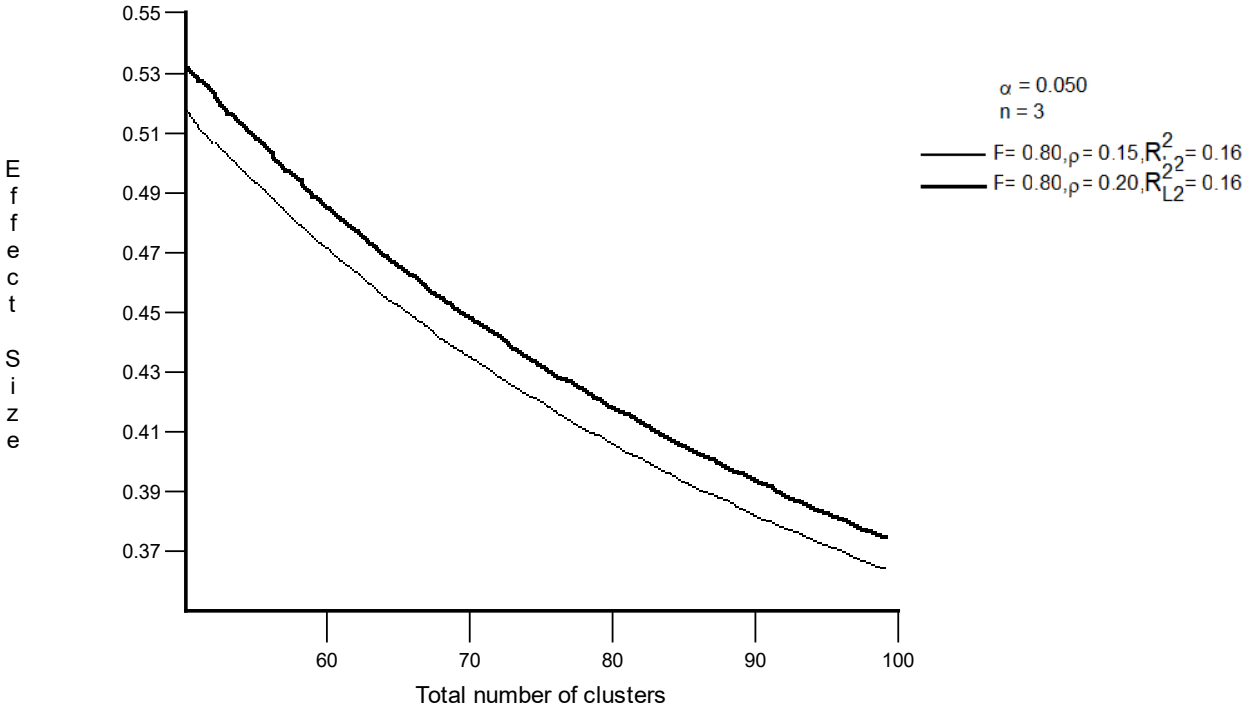


[4] DfE 2018: 11

[5] To our knowledge, there is currently no publicly available information on the likely correlation between pre- and post-test scores for our primary outcome measure, the early word recognition subscale from the York Assessment of Reading for Comprehension. The efficacy trial of the Nuffield Early Language Intervention (implemented with a similar age group) found that around 55% of the variation in post-test scores (a composite language score) was explained by the pre-test and pupil characteristics when using the full YARC (Sibieta et al., 2016). This would be higher due to the additional inclusion of pupil characteristics; we therefore use a lower estimate of 16 per cent for our assumptions.

[6] Previous EEF evaluations on early years have indicated that schools explain around 15 per cent of the variation in pupil attainment. For example, the efficacy trial of EasyPeasy found an ICC of 0.18 (Robinson Smith et al. 2019), whilst the efficacy trial of Family Skills reported an ICC of 0.15 at class level (Husain et al. 2018).

In the case of free school meals, DfE figures indicate that 14.9 per cent of nursery and primary school children in Inner and Outer London were known to be eligible for and claiming free school meals in January 2018. Allowing for attrition between pre- and post-testing, this equates to around three children per class. As a result, the MDE would be 0.37 standard deviations if the intra-class correlation was 0.15 and all other assumptions remained the same. The MDE would also be 0.38 standard deviations if the intra-class correlation was 0.20. Given the large MDE, it may be not be possible to discern whether the intervention has had a clear impact on this particular subset of pupils.



Outcome measures

Baseline measures

Two subscales from the York Assessment of Reading for Comprehension (YARC) Early Reading test[7] will be used as the baseline measure. The YARC Early Reading test is suitable for 4 to 7 year olds and covers four dimensions: sound isolation, sound deletion, letter sound knowledge and early word recognition. An overview of the four subscales are given below, including internal reliability scores and correlation with the Single Word Reading Test (SWRT)[8] as a measure of validity.

- The **Sound isolation** test measures phoneme isolation skills which are a component of phonemic awareness. Children hear a series of 12 nonsense words and are asked to identify either the first or the final sound in the word. The test's internal reliability using Cronbach's alpha is 0.88, and correlation with the SWRT was 0.62. This test is a measure of phonological awareness which is a longer term outcome/impact identified in the logic model.
- The **Sound Deletion** test measures phoneme deletion skills which are a component of phonemic awareness. Children hear a series of 12 words accompanied by a picture of what they represent and they are asked to repeat the word but 'take away' a sound from the word. The test's internal reliability

[7] Available from GL at https://www.gl-assessment.co.uk/products/york-assessment-of-reading-for-comprehension-yarc/

[8] The Single Word Reading Test (SWRT) is available from GL: https://www.gl-assessment.co.uk/products/single-word-reading-test-swrt/

using Cronbach's alpha is 0.93, and correlation with the SWRT was 0.76. If the Sound Isolation and Sound deletion scores are combined, this combined score has an internal reliability of 0.95 using Cronbach's alpha. This test is a measure of phonological awareness which is a longer term outcome/impact identified in the logic model.

- The **Letter Sound Knowledge** test measures alphabetic knowledge. Children are shown lower case letters and digraphs, one at a time, and are asked to say what sound the letters and digraphs make. The core test comprises 11 letters and 6 digraphs. The extended test comprises 26 letters and 6 digraphs. The Core test's internal reliability using Cronbach's alpha is 0.95. Its correlation with the SWRT was 0.55. This test is a measure of phonological awareness which is a longer term outcome/impact identified in the logic model.
- The **Early Word Recognition** test measures reading attainment in young readers. Children are asked to read 30 single words which are graded in difficulty. Half of the words have regular correspondence between the graphemes and phonemes, i.e. letter to sound mapping, and half are irregular. The test's internal reliability using Cronbach's alpha is 0.98, and correlation with the SWRT was 0.88. This test is a measure of overall literacy outcomes which are a longer term outcome/impact identified in the logic model.

To reduce testing time and burden on the school, only the early word recognition (measured on a scale of 0-30) and core letter sound knowledge (0-17) subscales will be used as a pre-test. These two subscales are most likely to be affected by the intervention, as the early word recognition subscale reflects children's ability to read, whilst letter sound knowledge is more appropriate than the other subscales given the timing of the pre-test.

These two tests will be used in combination as the baseline for the secondary outcomes, whilst the early word recognition subscale will be used as the baseline for the primary outcome. The baseline tests will be administered in November 2020.

Primary outcome

The primary outcome will be the early word recognition subscale score as the theory of change model suggests that this is where most of the impact will be seen. Qa Research will carry out the pre and post- tests as an independent test administrator and administrators will be blind to whether the school has been assigned to the treatment or control group.

Secondary outcomes

All four subscales from YARC will be used as a post-test, with a composite measure of Early Word Reading constructed by standardising the raw scores on each of the subscales and then combining them into a single metric.

To capture the differential impact of the Direct Mapping and Set for Variability strategies we will use an adapted version of the Mispronunciation Correction Test (MCT) developed by Tunmer and Chapman (2012) to assess the impact of Set for Variability. During the MCT, children are asked to play a game with a puppet that they are told sometimes says words wrong. The children are then presented with examples of irregular words that are incorrectly pronounced with a regularised pronunciation. The children are asked to 'correct' the puppet which requires them to consider other possible pronunciations of the word they have just heard, i.e. implement a set-for-variability strategy. The number of words that a child successfully 'corrects' is then used as a measure of their ability to use set-for-variability strategies. In their 2012 paper, Tunmer and Chapman deliver the test as two sessions conducted two weeks apart (one presenting the word in isolation and one where the words are embedded in sentences), but for this study, children will be tested once with the puppet

pronouncing the words embedded in sentences following the methodology used by Dyson et al. (2017).

The MCT will be administered at post-test only due to the limited time to carry out testing in the period prior to the delivery of the intervention and only for a subsample of 15 pupils per school. This will reduce the burden on schools and pupils and keep costs to a minimum. We will seek to explore the means by which SfV affects the primary outcome measure in order to isolate the impact of direct mapping and compare the differential impact of each strategy. Again, all the post-tests will be administered by Qa Research and testing will take place in June and July 2021.

In addition to the post-test outcome measures, we will use data from the Year 1 Phonics Screening check to explore the impact of the intervention on longer-term outcomes. This is discussed in further detail in the section on longitudinal follow-ups.

Compliance

Our primary analysis will capture the intention- to-treat effect and will not take into account non-attendance at training sessions, or schools ceasing delivery during the year. We will work with UCL to ensure the engagement of teachers, TAs and schools to minimise non-compliance, which will be examined through the implementation and process evaluation (IPE).

Whilst the focus will be on the impact of the intention-to-treat, we will also use a measure of compliance to determine how the intervention affects participants. The measure of compliance will be formed of at least one teacher/ TA's[9] attendance at all three training sessions and then based on an observation by a delivery team RA at the follow- up sessions, a rubric will be filled in examining how many strands of the content is present (Professor Savage has suggested two would need to be present to be compliant). We will develop this rubric in collaboration with the delivery team during the course of the pilot and test it and revise it if necessary for the main trial. The protocol will be updated to detail the confirmed rubric after the pilot.

Analysis

The analysis will use multi-level modelling, to take account of the trial's nested structure. The primary outcome will be measured at pupil level and the primary analysis will control for prior attainment, to increase statistical power and precision of the impact estimate (following EEF guidance).

Estimated impacts will be converted into Hedges' *g* effect size (1981) which uses the estimated total pooled standard deviation of the treatment and control groups. This provides a more conservative estimate of impact compared with using the within-school pooled standard deviation. Hedge's *g* effect sizes will be reported along with 95 per cent confidence intervals, as per EEF reporting guidelines and the analysis will explore whether impact estimates are statistically significant at the 5 per cent level or better.

We will conduct separate analysis of the subgroup of pupils eligible for Free School Meals (FSM), using the NPD indicator of whether the pupil has ever been eligible for FSM (EVERFSM_6_P) interacted with the treatment indicator. We will also seek to estimate the differential impact of the intervention on low-ability pupils by again including an interaction

[9] We will explore whether teacher or TA attendance or both is necessary during the pilot and update the protocol to confirm.

term in the model. Low-ability pupils will be defined as those who score less than the median on the pre-test (compared to those scoring over the median score on the pre- test) for the primary outcome only.[10] A subgroup analysis will also be used to explore the impact of the intervention in schools with existing good phonics practice. Where possible this will also estimate the differential impact of the intervention in schools with/without existing good practice by including an interaction term.

The delivery team will be asked to collect information on the percentage of pupils passing the phonics screening check administered in 2019 as a measure of existing good phonics practice. Provided this information is forthcoming, it could be used to identify schools with an above-average proportion of pupils passing the screening check. The protocol will be updated after the process of this has been confirmed.

82% of pupils passed the phonics screening check in 2019[11] and so if schools where the percentage of pupils passing the phonics screening check exceeds the national average are considered to provide good phonics teaching a high proportion of schools will fall into this category. The fact that a large proportion of schools are likely to have existing good practice makes it feasible to explore the impact of Flexible Phonics in the subset of schools with existing good practice. If numbers allow, we will also look at the differential impact of the intervention on schools with existing good practice compared to those without.

Whilst our preference would be to identify schools with good phonics practice based on their performance in the phonics screening check, if it proves impossible to gain full information on this measure as part of the recruitment process, we would use the publically available data on progress between Key Stage 1 and Key Stage 2 to identify good phonics practice.[12] Schools which made above-average progress in reading between Key Stage 1 and Key Stage 2 would be classified as having good phonics practice. The subgroup analysis would include an interaction term between the intervention dummy and the quality of existing practice to capture the added-value of the intervention in schools with existing good practice and indicate whether the intervention was more or less effective in schools that performed well prior to the trial[13].

If the analysis suggests that treatment and control groups differ on characteristics likely to be related to outcomes at baseline, we will explore whether the findings are robust to the inclusion of these additional characteristics (as a secondary model). We will also explore the impact of controlling for the full range of characteristics prior to treatment, alongside a simplified model based on differences in outcome between treatment and control groups. This will indicate how varying the model specification affects likelihood of detecting impact.

We will explore the impact of the two strategies using path analysis to identify the relative effectiveness of each, drawing on existing evidence on causal pathways to capture the differential impact of each alongside direct/indirect effects on the primary outcome. To

[10] As an alternative, we considered whether it would be feasible to explore the differential impact of Flexible Phonics throughout the ability distribution. Whilst in theory this could shed light on whether it is particularly effective for subsets of pupils, this would add to the complexity of the analysis and with expected sample sizes there is a risk that findings could be inconclusive or difficult to interpret. Existing evidence from Canada (Savage et al. 2018) indicates that Flexible Phonics is effective for low ability pupils and so it seems prudent to focus on ascertaining whether this is the case in the English context in the current study.

[11] As reported in Department for Education (2019).

[12] Available at: https://www.compare-school-performance.service.gov.uk/

[13] We have rejected the option of oversampling schools that do not have good phonics teaching to enhance the prospects of being able to explore differential impacts because of the risks this poses to recruiting sufficient schools for the trial.

assess the differential impact of DM and SfV, we will use MCT to measure SfV, and then multilevel generalised path analysis to disentangle the part of the intervention effect that can be ascribed separately to DM (and any other effect) and SfV. A full statistical analysis plan will be delivered that complies with EEF's guidance.

Longitudinal follow-ups

Additionally, we propose a delayed outcome measure. We have considered re-administering the post-intervention test during 2021/2022. However, the costs and time implied for schools make this impractical. Instead, we will use Year 1 phonics screening check data, which will become available in September 2022 to provide a delayed outcome measure at comparatively low-cost. The analysis will take a similar approach to that used in the primary analysis i.e. it will be based on the intention to treat.

The phonics screening check is a statutory assessment which takes place at the end of year 1 when pupils are typically aged 6. It confirms whether pupils have met the expected standard in phonic decoding. All state-funded schools with a year 1 cohort must administer the check. Teachers administer the check one-on-one with each pupil and record whether their response to each of the 40 words is correct. The words include 20 words and 20 pseudo-words. Pupils are told before reading each page of the test whether the words presented are real words or words for imaginary animals, i.e. pseudo-words. Each pupil is awarded a mark between 0 and 40, and in 2019, the threshold to determine whether a pupil had met the expected standard was 32. In 2022, the phonics screening checks are scheduled for the week commencing Monday 6 June. The Flexible Phonics intervention will run until June 2021 so there will be approximately 12 months between the end of the intervention and the phonics screening assessment.

The phonics screening check is an imperfect measure of the added benefit of the intervention as it measures the decoding of regular words, whereas the intervention focuses on the words that break the phonic rules, however it is the only freely available phonics measure that can be collected systematically across all schools and so provides an indication of longer term outcome in this area which may still be affected by the intervention.

Implementation and process evaluation

Research questions

The IPE will assess the eight key implementation dimensions set out below and identify moderating/contextual factors that influence impact and explain quantitative findings. It will gather evidence to inform any effectiveness trial.

Key research questions are:

Fidelity:

- IPE1: Are schools delivering the interventions and the trial as intended[14]?
- IPE2: Could the intervention be rolled-out on a larger scale so that the intervention is delivered as intended?
- IPE3: What adaptations would be required to roll-out the intervention on a larger scale and how might these affect the integrity of how the intervention is delivered?

[14] Key features of programme delivery will be agreed with the delivery team once the pilot has been completed and the protocol will be updated accordingly.

Dosage:

- IPE4: Do teachers and/or TAs teaching reception receive all intended training?
- IPE5: How often do participating teachers and TAs deliver Flexible Phonics strategies in phonics teaching?

Quality:

- IPE6: How well is initial training and follow-up support received by teachers, TAs and senior leadership at the school?
- IPE7: Is it necessary to conduct cascading training - has this been monitored/supported?
- IPE8: How effectively do teachers/TAs use Flexible Phonics strategies?
- IPE9: What facilitates/hinders effective implementation?
- IPE10: Would teachers/TAs find additional support helpful in maintaining quality - what and from whom?
- IPE11: Are there unintended or negative effects of the intervention?
- IPE12: What are TAs/teachers' perceived benefits and outcomes of the intervention?

Reach:

- IPE13: Do all intended pupils receive Flexible Phonics teaching?
- IPE14: Do some pupils receive more Flexible Phonics teaching than others?

Responsiveness:

- IPE15: Do teachers/TAs engage well?
- IPE16: Is the intervention acceptable and practicable in schools' contexts?
- IPE17: Do senior managers perceive the intervention as worthwhile/cost-effective?

Programme differentiation:

- IPE18: How does the intervention enhance/differ from existing phonics teaching?
- IPE19: Before Flexible Phonics implementation, what was business-as-usual and how was this embedded in wider approaches to reading?

Control group monitoring:

- IPE20: What phonics teaching and wider reading strategies are used in control schools?
- IPE21: Does the behaviour of control schools change during the trial?

Adaptation:

- IPE22: Have schools adapted the intervention - how and why?

Research methods

Drawing on the EEF guidance (Humphrey et al 2016, EEF, 2019a) we will use a multiphase mixed methods design involving:

- Two IDEA workshops and reviewing programme materials;

- Observations of one pilot training day, and review of UCL pilot reports.
- Observation of three online training half days for the main trial and two follow-up training sessions;
- Online surveys (baseline and post-treatment) of Reception teachers/other staff to gather evidence about business-as-usual and changes to practice;
- Case study visits to eight intervention schools towards the end of the programme to observe teaching, interview with Reception teachers and/or TAs who are involved in teaching phonics to the class participating in the study, literacy or early years leads and a senior leader;
- Interviews with UCL; and
- Analysis of data collected by UCL e.g. attendance and cost data.

In the **IDEA workshops**, our own and UCL's team will explore the intervention as part of an initial session shortly after set-up (October 2019) and then another after pre-trial development (in June/ July 2020). Building on the set-up meetings, we would: co-develop the TIDieR framework and theory of change; examine training/delivery materials; re-visit evidence about the interventions. This will lay a solid foundation for the evaluation and enable us to tackle key questions such as an appropriate compliance measure.

As part of the UCL pilot stage, we observed one training day. No further attendance was possible due to the covid-19 pandemic[15]. We will also read the two UCL pilot reports to learn more about the intervention and how it develops through the pilot to further develop the theory of change and prepare our observation, interview and survey materials for the main trial.

We will then **observe three half day training online sessions** on the virtual meeting platform such as Zoom for the main trial (instead of the original two in-person training sessions - 1.5 days in total) and two-follow up sessions (which may be virtual instead of in-person depending on the status of the covid-19 pandemic) to understand expectations for delivery and if things have changed from the pilot and how effective that has been, to underpin the IPE. Observations will help us to develop well-tailored research instruments (case study topic guides, post-intervention survey of teachers and TA) which will be quality assured by the lead on IPE. By observing all three online training half days, we will be able to observe all of the training material being delivered and also possible differences between training cohort groups (of which there will be six groups in total). The delivery team is doing this to have smaller training group sizes which makes the practical nature of the sessions (such as breakout groups and any interactive activities) easier to manage.

The survey of teachers and other key staff (across all schools) will be developed by IES using online survey software, SNAP, which allows completion on mobile devices. The baseline survey will capture usual- practice to randomisation, and information on broader approaches to teaching reading. We will use resources such as the Ofsted report on features of a good and outstanding Reception curriculum, which include characteristics of strong phonics teaching[16] and EEF guidance on improving literacy at Key Stage 1[17] to formulate questions. This will support the identification of value-added impact e.g. in combination with data on past reading attainment it will be possible to explore links between usual-practice and past performance. It will also help identify the extent to which results may be explained by control schools improving phonics teaching (due to compensatory

[15] We were due to also attend a follow- up session and two school support sessions.
[16] Ofsted, 2017
[17] EEF, 2017

rivalry/other drivers) or displacement of other literacy activities in treatment schools. The endline survey will repeat questions about phonics teaching and, for treatment schools, cover experiences of taking part, staff time and resources required (to inform the cost-per-pupil estimate). It will also include questions on adaptations made to the programme (beyond expected differentiation to meet the needs of individual pupils). We anticipate the surveys will be sent to approximately 230 teaching staff, including teachers and/or TAs, at least at both time points. The surveys will be census surveys sent to all teachers/TAs of participating classes and at endline we will review any differences between non-responders and those who have completed the survey.

Eight case studies will allow detailed qualitative exploration of delivery. The sample will be selected to include schools of different sizes/types, and potentially a mix of compliance. We will explore this selection process further during the pilot. If the baseline survey indicates differences in pre-intervention phonics teaching, this will be built into sampling. In each case study, we will first observe teachers/TAs using Flexible Phonics strategies and use these data to inform assessment of fidelity and shape questions for follow-up interviews. We will also interview a literacy/early years lead and senior leader. We anticipate an individual interview with senior leaders and will interview teachers and TAs separately to ensure open/honest discussions. Interviews will explore: training received; materials; workload/time requirements of Flexible Phonics; costs incurred; facilitators/barriers to implementation; adaptations and reasons (including views on how children with SEND/EAL respond); how Flexible Phonics compares to usual-practice; pupils' outcomes; and suggested improvements. Interviews will last around 45 minutes. We expect to interview about 3-4 participants per case study so up to a total of 32 participants.

We will also conduct **6 telephone interviews with UCL** towards the end of intervention delivery, including Professor Savage, the project manager and four research assistants. These will explore delivering training, school engagement and participation, and enablers/barriers to successful implementation of the programme. This should provide a rich picture of how training and support was delivered to schools across geographical areas, school types and pre-existing phonics programmes.

Finally, we will analyse data collected by UCL. We will work with them to specify additional data requirements during the pilot, which will cover teacher/TA attendance at training (which will be used for the compliance analysis) and also potentially school requests for additional support. In addition we will collect the observational data that the RAs collect at the follow-up visits which will also form part of the compliance analysis. This data would be analysed/triangulated with the evaluation data. We would ask UCL to inform us if schools withdraw, and will conduct short telephone interviews (up to nine) with such schools to explore reasons or via an email form. We will collect cost data from UCL to calculate the cost-per-pupil, including fees charged and length of training days/visits.

Analysis

Interviews will be digitally recorded with the agreement of participants and transcribed verbatim. We will analyse data using 'framework', drawing themes and messages from an analysis of interview transcripts, observations of training, and other materials collected by evaluation and project teams as a pragmatic cost- effective approach for this amount of qualitative data. Data will be collected using the methods described in the table below and analysed according to the research questions listed.

Framework is an excel-based qualitative analysis tool that ensures that the analytical process and interpretations from it are grounded in the data and tailored to the research

questions. It was designed to ensure a systematic and consistent treatment of all units of data (e.g. transcripts of interviews). It also allows for the analytical framework to be refined and modified in the early stages of its use.

The context of the information is retained and the page of the transcript from which it comes is noted, so that it is possible to return to a transcript to explore a point in more detail or to extract text for verbatim quotations.

Framework allows full within case analysis (looking in detail at each individual case) and between case analysis (comparing individual cases and groups of cases), and it is the ability to interrogate data at both these levels that adds real richness and depth to the analysis and interpretation. Organising the data in this way allows us to compare the full range of experiences and accounts and patterns across different groups of people.

Observations and themes identified in the qualitative data through the fieldwork can then be compared with quantitative data gathered such as survey findings, training attendance, etc, to test whether perceptions are reflected across the settings overall. Triangulating rich qualitative and quantitative data will allow us to interrogate the mechanisms proposed in the logic model, understand factors contributing to outcomes or identify barriers/enablers.

Table 3: IPE methods overview *(adapt as necessary)*

Data collection methods	Participants/ data sources (type, number)	Data analysis methods	Research questions addressed	Implementation/ logic model relevance
Two IDEA workshops	Evaluation team and delivery team	Theory of Change	IPE 5, 15.	Theory of Change, Inputs, Activities
Reviewing intervention materials	Training materials, support materials	Literature review, thematic analysis	IPE 2, 3, 16, 18.	Inputs, Activities
Observations of pilot training day	Delivery team (Prof. Savage) and Reception teachers/TAs from pilot schools	Observation framework, Theory of Change	IPE 2, 7, 10, 18.	Input, Activities, Outputs
Observation of three online training half days for the main trial and two follow- up training sessions	Delivery team (Prof. Savage and RAs) and Reception teachers/TAs from half of schools receiving the intervention (~25-35 schools)	Observation framework	IPE 2, 4, 7, 10, 15, 18.	Input, Activities, Outputs

Online surveys (baseline and post-treatment) of Reception teachers/other staff	Reception teachers/ TAs from all schools participating in the study (<=120)	Descriptive: frequencies, cross-tabs, t-tests, ANOVA, regression	IPE 7, 9, 10, 12, 13, 15, 17, 19-21.	Activities, Outputs, Outcomes, Enabling factors
Case study including observations and interviews	8 intervention schools, 8 or more Reception teachers/TAs , Up to 8 literacy or early years leads, up to 8 senior leaders;	Teaching observation framework, Extraction framework	IPE 1-3, 4-5, 6-12, 13-14, 15-17, 18-19, 22.	Inputs, Activities, Outputs, Outcomes, Enabling factors
Interviews with UCL	Intervention designer/trainer: Prof. Savage, RAs undertaking support visits	Extraction framework	IPE 1-4, 7, 15, 18.	Inputs, Activities, Enabling factors
Analysis of intervention data collected by UCL	Training attendance, summary of compliance, satisfaction surveys, and cost data	Thematic analyses, Descriptives :frequencies , cross-tabs, average cost per child.	IPE 2-4, 7, 13, 15.	Inputs, Activities, Outputs

Cost evaluation

Cost data will be collected directly from the delivery team on their own costs for running the training and follow- up sessions and materials provided and then from schools cost and time data will be collected during the post survey on training (and whether or not cover was required) and other time taken for preparing for the lessons/ reading materials etc. .

A cost per pupil over three years will be calculated using the new EEF cost guidance (EEF 2019b).

Ethics and registration

IES will apply for ethics approval through the internal IES system once the recruitment materials are finalised in January 2020. Schools will sign a MOU to agree to take part in the project and an information pack of details of the project. Teachers/ TAs will also receive an information sheet detailing the trial and data sharing and will have the chance to opt out of taking part in the IPE when they receive the invitation to complete the survey/ take part in an interview. Schools are all also being invited to webinars in July with the UCL, IES and Qa teams to hear more about the project and ask any questions they have.

Parents will receive an information sheet detailing the trial and data sharing and will have a chance to withdraw their child's data from being shared with the evaluation team at the start of the trial by telling the school. After randomisation the parents can contact IES directly to have their child's data removed from the data stored by the evaluation team.

The trial will be registered at www.controlled-trials.com once the protocol has been finalised and will be updated when necessary.

Data protection

We take seriously the ethical issues raised by this research, including burden on schools and pupils. We also recognise the utmost importance of data protection and are fully committed to complying with the Data Protection Act 2018 legislation and will carry out a Data Protection Impact Assessment for this project.

All participants interviewed will sign a consent form to indicate that they have understood the research aims, agree to the interview being recorded/transcribed, and will be given assurance of anonymity. Schools will sign a MOU identifying the requirements of the project and how the data will be used, shared and stored. Parents will receive a letter explaining the trial and how they can withdraw their child's data.

We will develop a privacy notice explaining how information collected will be used and stored, and to communicate to participants their right to withdraw from data processing. This will be available online, with the link provided in letters/briefings. We will also develop a data-sharing agreement between IES, UCL and EEF stating data to be shared by whom, how and why, to ensure full data-security.

Our approach involves personal data collection including pupil name, date of birth, gender and Unique Pupil Number (UPN). IES will access and link this pupil data to background and school data held on the National Pupil Database (NPD). The NPD data to be requested will include whether or not the pupil is eligible for Free School Meals (FSM) as well as their gender and their phonics score in Year 1. IES will match the above pupil data to data on pupil outcomes. This will include data from questionnaires and assessments administered as part of the project including a standard assessment of literacy skills and a measure of mispronunciation correction as well as data on outcomes available through the NPD.

IES' legal basis for processing personal data is 'legitimate interests'. The evaluation of Flexible Phonics fulfils one of IES' core business purposes (undertaking research, evaluation and information activities) and is therefore in our legitimate interest, that processing personal information is necessary for the conduct of the evaluation. IES is required to process data for the following tasks as part of the evaluation:

- To conduct the randomisation (i.e. to randomly allocate schools to receive or not receive the intervention)
- To match data received from schools to NPD data and outcome data
- To contact teachers about participating in interviews and surveys
- To instruct and liaise with independent test administrators
- To evaluate the impact and effectiveness of the programme and prepare a report about the project

No special category data will be collected or processed as part of this evaluation.

For the purposes of conducting the evaluation to assess the impact of Flexible Phonics, IES and IOE will both become data controllers of personal data of school staff and pupils

obtained from schools and other sources such as the National Pupil Database. Personal data may be shared with trusted processors such as test administrators and transcribers as well as members of the delivery and evaluation teams solely for the purposes of proper delivery, management and evaluation of the project. At the end of the project, data will be submitted to the EEF's data archive. At this point, EEF will become a data controller and the archive manager will be a data processor.

IES will securely delete all personal data within six months of the project finishing, i.e. once the final draft of the addendum report has been submitted and the trial data has been submitted to the EEF archive. UCL IOE will keep the data for five years.

Personnel

Delivery team

Professor Rob Savage- UCL IOE- developer of the Flexible Phonics intervention

Amy Fox - UCL IOE -Project Manager

Research Assistants to be recruited December 2020

Evaluation team

Dr Anneka Dawson Co- Principal investigator of the evaluation. Anneka will lead on the implementation and process evaluation (IPE) and overseeing the assessments and quality assuring materials.

Dr Helen Gray- IES. Co- Principal investigator of the evaluation. Helen will lead on the impact assessment.

Dr Clare Huxley- IES. Project Manager will be responsible for managing all research activity, liaising with UCL, drafting research tools and coordinating members of the research team.

Dr Susie Bamford- IES associate. Susie will support the impact evaluation analysis.

Dr Dafni Papoutsaki, IES. Dafni will also support the impact evaluation analysis.

Kate Alexander, IES. Kate will support the IPE.

Julie Vanderleyden, IES. Julie will support the IPE.

Risks

Issue/risk	Action to address issue/reduce risk	Likelihood	Impact
Delays to the timetable	Experienced trial lead and project manager; close communication between delivery team and evaluators particularly during the autumn term, which is a critical point in timetable. Careful selection of a pre-test	Medium-High	Medium

Issue/risk	Action to address issue/reduce risk	Likelihood	Impact
	measure to ensure it is as practicable as possible within the narrow timeframe available. Clear communications with schools at the recruitment stage about the timeframe for the pre- and post-test		
Covid-19 causing school closures or other disruption	Some of the pilot observations were cancelled as the pilot work changed to include a survey rather than in-person sessions. The training has been rescheduled to three half days in a virtual meeting platform rather than 1.5 in-person days and therefore the observations will also take place on this platform. UCL surveyed teachers to gauge interest in carrying on with the testing and training in person or virtually and testing will go ahead as planned. Further updates may be necessary if there is further disruption to schools which may mean follow- up sessions by UCL may need to be virtual and the corresponding observations and interviews may also need to be virtual. This will be reviewed in early 2021 and the protocol updated if necessary.	Medium	Medium
Pre- or post- test data completion rates are low	We will develop clear communications about testing requirements at the recruitment stage. We will work closely with the team recruiting schools to make sure they feel well-briefed and comfortable explaining the process in detail, listening to and addressing schools concerns. This will include providing information about the skills and experience of test administrators. We will work with the test administrator organisation and the UCL team to have school details as early as possible to plan testing at both time points. Mop- up visits will be completed by the test administrator wherever possible when there is absence to maximise the number of pupils tested. Progress will be monitored carefully	Medium	High

Issue/risk	Action to address issue/reduce risk	Likelihood	Impact
	throughout the testing period so that action can be taken swiftly if required.		
Reluctance of schools to participate in case study research or survey research	Schools will have their responsibilities clearly laid out in a Memorandum of Understanding (MoU). It is likely that not all TAs and teachers will be aware of the detail of the MoU so we will produce early on a clear and concise research briefing for teachers, which explains the requirements and timings of the evaluation activity and provide a point of contact for questions which we will include in an information pack. All schools will be invited to take part in a webinar for further information and to ask questions. We will work to reduce burdens on participants as much as possible with short surveys and interview times. A broad team means we can be flexible about dates for case studies. As a contingency, if it is not possible to set up face- to- face visits at schools, we would consider conducting depth telephone interviews with school staff instead but this would be discussed with EEF and used only as a last resort.	Low	High
Schools drop out	Evaluators have limited ability to affect participant numbers other than explore reasons for low participation, identify good practice in maximising participant engagement and share this across schools. We can work to maximise research participation among those taking part, including trying to keep control schools engaged by being in relatively frequent contact.	Low	Medium
Delays to receiving NPD data for addendum report	Submit early application; experienced researcher to oversee application to ensure no delays due to incomplete application	Medium	Low

Timeline

Table 4: Timeline

Dates	Activity	Staff responsible/ leading
June-Oct 2019	Set up meetings and first IDEA workshop	Delivery team and evaluation team
Sep 2019- Feb 2020	Recruitment of pilot schools	Delivery team
Dec 2019- July2020	Recruitment of trial schools	Delivery team with support from evaluation team
Oct 2019- June 2020	Pre-trial development of programme Observation of pilot training session	Delivery team Evaluation team
July 2020	Webinars - school Information session for participating schools.	Delivery Team with support from Evaluation and Assessment Team
Aug- Sep 2020	Second IDEA workshop	Delivery team and evaluation team
Sep- Dec 2020	Collection of pupil data Collection of pre-test data Business as usual survey of teachers/ TAs Randomisation	Delivery team and evaluation team (overseeing Qa Research test administrator)
Jan- June 2021	School training days by end Feb and school visits by the end of March 2021 Observation of school training days Schools deliver Flexible Phonics	Delivery team and evaluation team
Apr- Jun 2021	Collection of data from delivery team School case studies (observation, senior leader, teacher/TA and literacy lead/ early years lead/SENCO interviews)	Evaluation team
June- July 2021	Administration of post –test assessments	Evaluation team (overseeing Qa Research test administrator)
June- July 2021	Post- intervention survey of teachers/ TAs	Evaluation team
Autumn 2021	Analysis of project and evaluation data	Evaluation team
Dec 2021- Jan 2022	Evaluation report writing	Evaluation team
Jan 31st 2022	First draft of evaluation report	Evaluation team
Sept 2022	Obtain NPD data for Year 1 phonics	Evaluation team
Autumn 2022	Analysis of Year 1 phonics and evaluation data for addendum report	Evaluation team
Jan- Feb 2023	Addendum report writing	Evaluation team

Dates	Activity	Staff responsible/ leading
Jan 31st 2023	First draft of addendum report	Evaluation team

References

Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2008). Meta-Analysis of the effects of early education interventions on cognitive and social development. Teachers College Record, 112.3: pp. 579–620.

Chen, V., & Savage, R. S. (2014). Evidence for a simplicity principle: Teaching common complex grapheme-phonemes improves reading and motivation in at-risk readers. Journal of Research in Reading, 37, 196–214. doi:10.1111/1467-9817.12022

Department for Education (2010), 'The importance of teaching: the schools white paper 2010', White Paper, November 2010.

Department for Education. (2011) *New phonics check will identify thousands of children needing extra reading help* [Online]. Available from: https://www.gov.uk/government/news/new-phonics-check-will-identify-thousands-of-children-needing-extra-reading-help [Accessed 18.December 2019].

Department for Education. (2014) *Phonics: Choosing a programme* [Online]. Available from: https://www.gov.uk/government/collections/phonics-choosing-a-programme [Accessed: 18 December 2019]

Department for Education. (2017), 'Statutory framework for the early years foundation stage', March 2017.

Department for Education (2018) 'Schools, pupils and their characteristics: January 2018'. Statistical Publication. 28 June 2018.

Department for Education (2019) Phonics screening check and key stage 1 assessments in England, 2019. 26 September 2019. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/851296/Phonics_screening_check_and_key_stage_1_assessments_in_England_2019.pdf

Dyson, H., Best, W., Solity, J., & Hulme, C. (2017). Training mispronunciation correction and word meanings improves children's ability to learn to read words. Scientific Studies of Reading, 21, 392–407. doi:10.1080/10888438.2017.1315424

EEF (2019a). Implementation and process evaluation guidance for EEF evaluations. The Education Endowment Foundation, London. Available at:

https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/IPE_guidance.pdf

EEF (2019b). Cost evaluation guidance for EEF evaluations. The Education Endowment Foundation, London. Available at:

https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/Cost_Evaluation_Guidance_2019.12.11.pdf

Elbro, C., De Jong, P., Houter, D., & Nielsen, A.-M. (2012). From spelling pronunciation to lexical access: A second step in word decoding. Scientific Studies of Reading., 16, 341–359. doi:10.1080/10888438.2013.8119356

Elbro, C., & De Jong, P. F. (2017). Orthographic learning is verbal learning. The role of spelling pronunciations. In K. Cain, D. Compton, & R. Parrila (Eds.), Theories of reading development (pp. 169–190). Amsterdam, the Netherlands: John Benjamins.

Galuschka, K., Ise, E., Krick, K. & Schulte-Körne, G. (2014). Effectiveness of Treatment Approaches for Children and Adolescents with Reading Disabilities: A Meta-Analysis of Randomized Controlled Trials PLoS ONE 9(2): e89900.

Gibson, E. J. (1965). Learning to read: Experimental psychologists examine the process by which a fundamental intellectual skill is acquired. Science, 148, 1066–1072. doi:10.1126/science.148.3673.1066

Hatcher, P. J., Hulme, C., & Ellis, A. W. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. Child Development, 65, 41–57. doi:10.1111/j.1467-8624.1994.tb00733.x

Hatcher, P. J., Hulme, C., Miles, J. N. V., Carroll, J. M., Hatcher, J., Gibbs, S., . . . Snowling, M. J. (2006). Efficacy of small group reading intervention for beginning readers with reading-delay: A randomised trial. Journal of Child Psychology and Psychiatry, 47, 820–827. doi:10.1111/j.1469-7610.2005.01559.x

Hatcher, P. J., Hulme, C., & Snowling, M. J. (2004). Explicit phoneme training with phonic reading instruction helps young children at risk of reading failure. Journal of Child Psychology and Psychiatry, 45, 338–358. doi:10.1111/j.1469-7610.2004.00225.x

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R & Kerr, K. (2016). Implementation and process evaluation (IPE) for interventions in education settings: An introductory handbook. The Education Endowment Foundation, London. Available at: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/IPE_Guidance_Final.pdf

Husain, F., Wishart, R., Marshall, L., Frankenberg, S., Bussard, L., Chidley, S., Hudson, R., Votjkova, M. and Morris, S. (2018) Family Skills: Evaluation report and executive summary. The Education Endowment Foundation, London. Available at: https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Family_Skills.pdf

Kearns, D., Rogers, H. J., Koriakin, T., & Al Ghanem, R. (2016). Semantic and phonological ability to adjust decoding: A unique correlate of word reading skill? Scientific Studies of Reading, 20, 455–470. doi:10.1080/10888438.2016.1217865

Lovett, M. W., Lacerenza, L., Steinbach, K. A., & De Palma, M. (2014). Development and evaluation of a research based intervention program for children and adolescents with reading risabilities. Perspectives on Language and Literacy, 40, 21–29.

Robinson Smith, L., Menzies, V., Cramman, H., Wang, Y., Fairhurst, C., Hallett, S., Beckmann, N., Merrell, C., Torgerson, C., Stothard, S. and Siddiqui, N. (2019) EasyPeasy: Learning through play. Evaluation report. The Education Endowment Foundation, London. Available at: https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/EasyPeasy.pdf

Savage, R. S., Carless, S., & Ferraro, V. (2007). Predicting curriculum and test performance at age 11 years from pupil background, baseline skills and phonological awareness at age 5 years. Journal of Child Psychology and Psychiatry, 48, 732–739. doi:10.1111/j.1469-7610.2007.01746.x

Savage, R., Georgiou, G., Parrila, R. & Maiorino, K. (2018) Preventative Reading Interventions Teaching Direct Mapping of Graphemes in Texts and Set-for-Variability Aid At-Risk Learners, Scientific Studies of Reading, 22:3, 225-247, DOI:10.1080/10888438.2018.1427753

Savage, R. S., & Stuart, M. (2001). Orthographic analogies and early reading: Explorations of performance and variation in two transfer tasks. Reading and Writing, 14, 571–598. doi:10.1023/A:1012052631557

Savage, R. S., & Stuart, M. (2006). A developmental model of reading acquisition based upon early scaffolding errors and subsequent vowel inferences. Educational Psychology, 26, 33–53. doi:10.1080/01443410500340983

Shapiro, L. R., & Solity, J. (2008). Delivering phonological and phonics training within whole-class teaching. British Journal of Educational Psychology, 78, 597–620.

Sibieta, L., Kotecha, M. and Skipp, A.(2016) *Nuffield Early Language Intervention. Evaluation Report and Executive Summary*. The Education Endowment Foundation, London.

Steacy, L. M., Elleman, A. M., Lovett, M. W., & Compton, D. L. (2016). Exploring differential effects across two decoding treatments on item-level transfer in children with significant word reading difficulties: A new approach for testing intervention elements. Scientific Studies of Reading, 20, 283–295. doi:10.1080/10888438.2016.1178267

Torgerson, C., Brooks, G., Gascoine, L. and Higgins, S. (2019). 'Phonics: Reading policy and the evidence of effectiveness from a systematic "tertiary" review.' Research Papers in Education, 34, 2, 208-238.

http://www.tandfonline.com/doi/full/10.1080/02671522.2017.1420816

Torgerson, C., Brooks, G., & Hall, J. (2006). A Systematic Review of the Research Literature on the Use of Phonics in the Teaching of Reading and Spelling. Department for Education and Skills.

Tunmer, W. E., & Chapman, J. W. (2012). Does set for variability mediate the influence of vocabulary knowledge on the development of word recognition skills? Scientific Studies of Reading, 16, 122–140. doi:10.1080/10888438.2010.542527

Venezky, R. L. (1999). The American way of spelling: The structure and origins of American English orthography. New York, NY: Guilford Press.

Zipke, M. (2016). The importance of flexibility of pronunciation in learning to decode: A training study in set for variability. First Language, 36, 71–86.