

# Using Research Tools to Improve Language in the Early Years

Behavioural Insights Team and National Institute of Economic and Social Research

Michael Sanders



## Evaluation Summary

<b>Age range</b>	Nursery and Reception Year (ages 3 to 5)
<b>Number of pupils</b>	2,880 (24 children in 120 schools)
<b>Number of schools</b>	120 schools
<b>Design</b>	School-level cluster randomised control trial
<b>Primary Outcome</b>	Composite language skill score

## Background

### Intervention

This intervention aims to improve children's language and social-behavioural outcomes, by training nursery and reception teachers to use established research tools called the Early Childhood Environment Rating Scales (ERS). These scales are highly regarded research tools used for assessing the quality of early years settings, and have been shown in existing research studies to predict children's development in the early years and beyond. The rating scales provide a framework for assessing various elements of early years practice including language and reasoning, adult-child interactions, activities and care routines. Three of the ERS will be used: the ECERS-3, ECERS-E and SSTEW.

The model that will be tested in this programme comprises five days of training delivered to early years teachers over five months, with a follow-up day three months later. Practitioners are taught the principles of using rating scales to improve practice in their settings, and how to support children's development through evidence based practice. The programme is structured so that there is approximately one day of training per month, allowing time in between sessions for teachers to use the rating scales in their own settings, reflect on their practice, apply evidence-based pedagogical strategies to develop areas identified as needing improvement and involve other staff. Teachers will also receive mentoring to help them apply the training in practice. Mentoring support will be delivered through a mix of face-to-face and distance (skype) sessions, with approximately 3.5 days allocated per school (plus some contingency). Mentors will provide support in translating new learning into practice, engaging other staff (e.g. teaching assistants) in the change process and providing feedback through observation using the environment rating scales. Teachers and other staff within the school will also have access to online resources, including the DVD clips used during the training and a Pinterest board.

The intervention will be developed and delivered by a team from Oxford University, University College London (UCL) Institute of Education, and A+ Education Ltd. Schools will be recruited from the West Midlands, Liverpool and Manchester areas.

## Significance

There is robust evidence demonstrating the efficacy of interventions targeted during early childhood in improving both cognitive and non-cognitive abilities. In the cognitive space, early intervention programmes have been linked to improved measures of school readiness, academic attainment and school progression (Anderson et al, 2003). Participation in such programmes has also been associated with positive outcomes beyond school years, including lower rates of interaction with the criminal justice system, reduced likelihood of teenage pregnancy, reduced welfare dependency, and lower prevalence of risk factors for cardiovascular and metabolic diseases (Campbell et al, 2014; Currie, 2001; Gorey, 2001). Additionally, early interventions can provide a large return on investment, particularly compared to those occurring in later years, such as reducing pupil-teacher ratios (Heckman and Masterov, 2007; Heckman, 2012).

In the preschool context, several studies have investigated the impact of training teaching staff in new practices or curricula US studies have shown that training and supporting preschool teachers in classroom management strategies can reduce internalising and externalising behaviour problems in children and reduce signs of teacher-reported social withdrawal (Raver et al., 2009). Similarly, training preschool teachers in curricula designed to increase child social-emotional competence has been linked to improved child emotional knowledge skills, and parent and teacher-reported measures of social competence (Domitrovich, Cortes & Greenberg, 2007; Webster-Stratton, Reid, & Stoolmiller, 2008), as well as improved self-regulation (Webster-Stratton, Reid, & Stoolmiller, 2008).

Studies have also demonstrated the effectiveness of interventions in the early years for language development (Springate et al, 2008, provides a review). The Nuffield Early Language Intervention, comprising staff training (principally for teaching assistants), lesson plans and materials, was found to have a positive impact on children's language skills (Sibieta et al., 2016). This intervention targeted children with poor spoken language skills, however, there is also evidence that non-targeted programmes can be effective for improving language outcomes (Springate et al., 2008).

Of particular relevance to the present study, there is evidence from the US that promoting the capacity of teachers to use research-informed tools to improve their practice can advance both child social-emotional and cognitive development. The 'Research Based, Developmentally Informed' (REDI) cluster randomised control trial equipped Head Start preschool teachers with a portfolio of lesson plans and enrichment activities. Teachers also received substantial training in all materials and guides provided, and were supported by weekly meetings with a mentor. One-year after the trial began children who received the intervention had progressed significantly further in measures of vocabulary, emergent literacy and socio-emotional development (Bierman et al., 2008). A follow-up study tracking the children into kindergarten (one-year after the first study concluded) found evidence for a sustained intervention effect across most measures of socio-emotional development, but only detected a continuing significant effect in one measure of language development (phonemic decoding) (Bierman et al., 2014).

This project tests whether providing teachers with training in using the environment rating scales to promote quality practices in early years provision impacts upon children's language and social-

emotional development. These scales provide a framework through which the early learning environment can be assessed, focusing on child-centred pedagogy. The longitudinal Effective Provision of Pre-school, Primary and Secondary Education (EPPSE) study found that observed ratings of quality from the ECERS, controlling for other contributing factors, were associated with improved attainment in Maths and English, as well as better social outcomes, through to the end of Key Stage 2 (Sylva et al., 2008). Even at age 16, attending a higher quality pre-school (as measured by ECERS) was associated with better performance at GCSE, as well as improved self-regulation and pro-social behaviour (Sylva et al., 2014). Analysis of a sub-study of early years settings attended by children in the Millennium Cohort Study also showed a positive association between higher ECERS scores and improved language outcomes (Hopkin, Stokes and Wilkinson, 2010). The current evaluation aims to determine if training nursery and reception staff to better understand child development and curriculum, and to use ERS to evaluate and improve their practice, will lead to improved attainment in children's language and social development.

## Methods

### Research questions

The primary objective of this evaluation is to test whether the ERS training and mentoring intervention improves language development over the nursery and reception years of schooling, as assessed by performance in a battery of language skill assessments. These assessments consist of:

- **British Picture Vocabulary Scale (BPVS):** a measure of vocabulary.
- **Renfrew Action Picture Test (APT):** a measure of expressive language (both information and grammar).
- **Clinical Evaluation of Language Fundamentals (CELF) Preschool 2 UK - Sentence Structure subtest:** a measure of comprehension.

The evaluation will also address the following questions:

- What is the impact of the programme on children's social-behavioural outcomes, as measured by the Adaptive Social Behaviour Inventory (ASBI) score?
- Does the programme impact on the quality of provision for language and social-behavioural development as measured by a composite of items from the ECERS-3, ECERS-E and SSTEW - collectively the 'Environment Rating Scales' (ERS)?
- What is the impact of the programme on children's language and social outcomes for pupils who move to a reception class with a teacher who did not directly receive ERS training and mentoring?
- What is the impact of the programme on children's language and social outcomes for pupils eligible for free school meals?

- What is the impact of the programme on children’s language and social outcomes for pupils for whom English is an Additional Language?
- What is the impact of the programme on children’s language and social outcomes for pupils who have language difficulties at the start of the trial?
- Does the programme show differences in impact on children’s language and social outcomes according to children’s gender?

## **Design**

This will be a cluster randomised controlled trial. Randomisation will take place at school level. The trial aims to recruit 120 primary schools with nursery and reception classes, with schools randomly allocated to either the treatment arm (who will receive the programme) or the control group. Thus there will be two approximately equal arms of the trial. Schools in the control group will be expected to continue with ‘business as usual’, and will be offered the opportunity to take part in the programme following the completion of the study (August 2018), or a payment of £1,000.

## **Randomisation**

Randomisation will follow recruitment of schools, including the signing of Memorandums of Understanding (MoUs) and baseline data collection. Randomisation will be stratified on the basis of school-level characteristics to ensure balance between treatment and control groups (which will be of equal or near-equal size). This will be conducted using data analysis and statistical software Stata. The randomisation will follow a two-stage process:

1. The schools will be stratified on the basis of FSM students (split across the median sample proportion), local authority<sup>1</sup> and Key Stage 1 (KS1) English reading attainment (split across the median sample result).
2. A random number will be generated within each block and the subsamples split into two groups to ensure that school FSM proportion, KS1 performance and local authority are balanced across trial arms. We will use a combination of EduBase, the Department for Education’s Performance Tables, and school reported data to determine the blocking characteristics.

As noted above, in addition to stratifying on the basis of (grouped) local authority areas, we will also stratify on the proportion of FSM students in the school and school-level KS1 performance. Both characteristics are likely to be correlated with the outcomes measured in this trial. FSM is associated with lower attainment on average; stratifying on this variable may also help in ensuring sufficient sample sizes to allow subgroup analysis of the FSM group. In addition we stratify on KS1 attainment as a means of proxying school performance, also likely to correlate with the outcomes measured in the study. In a sample of the size planned for this trial, the correlation between the stratifying variable and the outcome variable only needs to be small in order to offset the loss of degrees of freedom.

---

<sup>1</sup> Local Authorities with fewer than 15 schools recruited to the trial will be pooled with their nearest geographic neighbour prior to stratification occurring.

Randomisation will take place once schools have signed the Memorandum of Understanding, returned parental consent forms and baseline data collection is complete. Schools will have their trial arm allocation communicated to them once baseline outcome measures assessments have been collected. The Stata code used for the randomisation will be recorded in the final report.

## **Participants**

The trial will be conducted across schools in the Liverpool Manchester and West Midlands areas. These areas have been chosen as they possess above average proportions of students from disadvantaged backgrounds (as measured by the proportion of neighbourhoods in the top 20 per cent of areas in the Index of Multiple Deprivation) and/or below average results at age 5 for communication and language (30% worst performing authorities for the proportion of children reaching expected level of development for communication and language in their EYFS profile). Some areas may be included in the trial on the basis of one of these two characteristics. The target number of schools for this intervention is 120.

Prior to recruitment, schools will submit an expression of interest form to A+ Education Ltd, which will include information that will allow eligibility to be determined. The project overview document circulated to schools will detail the eligibility criteria and provide further information about the study. Schools will be required to sign an MOU before enrolment into the study, agreeing to partake in the activities required for the intervention and evaluation to be carried out.

The eligibility criteria for schools include:

- Participating schools should be located in one of the study areas, and be a one or two form entry state primary school with a nursery class. Three or Four form entry schools will only be accepted where they agree to channel nursery children who have completed a baseline assessment into a reception class led by a participating teacher (defined as one nominated during the EOI process). This applies to both control and treatment schools. That is, in control group schools nursery children should move to a reception class led by a teacher who would have received the intervention had they been assigned to the treatment condition.
- One nursery and one to two reception teachers (with three reception teachers required where the school is three form or more entry) agree to attend the ERS training and engage with mentoring if allocated to the treatment group;
- Schools should not have not previously accessed training by A+ Education Ltd which is substantially similar to that being provided via the current intervention, received substantial support from their local authority using rating scales such as the Environment Rating Scales (ECERS and others) or used such tools themselves on a regular basis;
- If allocated to the control group, that schools continue with 'business as usual' for the duration of the trial;
- A completed MOU;

- Consent to participate in the study – including the collection of outcome measures in summer 2018 – regardless of which trial arm they are assigned to;
- Agreement to collect opt-in consent from the parents of children involved in the study, and the provision of both school and pupil level data.
- Agreement to allow time for each assessment phase and liaise with the evaluation team to find appropriate dates and times for assessments to take place; and
- Agreement that teachers in both trial arms complete a survey at the end of the trial period, and attend an interview with evaluation staff if requested.

Schools will be informed of their trial arm allocation after baseline outcome measures are collected.

Priority will be given to schools with a higher proportion of pupils eligible for FSM. Because of this, and the location of the study areas, the sample is expected to include a higher proportion of disadvantaged schools. There will not be specific eligibility criteria for students – all pupils in the nursery classes of participating schools will be eligible.

### **Outcome Measures**

The primary outcome measure will be a composite language skill score. This measure will draw on the results reported in three language assessments:

- **British Picture Vocabulary Scale (BPVS):** A one-to-one test that assesses a child's receptive vocabulary. For each question, the test administrator says a word and the child responds by selecting a picture from four options that best illustrates the word's meaning.
- **Renfrew Action Picture Test (APT):** In this test, the child is asked to describe the actions shown in a set of pictures. Two scores are recorded, one for the level of information they provide (for example nouns and verbs) and one for the grammar they use (such as use of tenses). Both information and grammar scores will be incorporated into the composite measure.
- **Clinical Evaluation of Language Fundamentals (CELF) Preschool 2 UK - Sentence Structure:** This subtest provides information about how a child understands spoken language. This is achieved by asking the child to interpret spoken sentences of increasing length and complexity by pointing to the picture that illustrates a given sentence.

To arrive at a composite language skill score we will standardise each of the components (including both APT scores) to have a mean of zero and standard deviation of one. These will then be added together to create composite measures and re-standardised. As such, the four language measure scores will be equally weighted in the composite language skill score.

All tests will be administered and scored by research assistants with an academic background in speech therapy or psychology. They will be recruited by BIT and given training in the language assessments used by an experienced child psychologist prior to visiting schools. Research assistants will be blind to the trial arm allocation of schools they visit. Tests will be conducted at two intervals during the course of the trial:

- **Pre-test:** this will be conducted during October-November 2016, prior to schools being informed of their trial arm assignment.
- **Post-test:** this will be undertaken during June-July 2018.

Individual scores for each of the language measures will also be reported as secondary outcome measures.

Additionally, during the pre-test outcome measure collection the evaluation team will trial the administration and scoring of the Renfrew Bus Story Test, a measure of narrative language skill, in one school. This will be done to gauge the approximate cost of collecting scores for this measure. If the cost of collection can be financed by an external partner, the evaluation team will also facilitate collection of mean-utterance-per-sentence for this test, and report this measure in our secondary analysis.

An additional secondary outcome will be social-behavioural development as measured by the Adaptive Social Behavior Inventory (ASBI) total score. The ASBI is a questionnaire that will be completed by class teachers for each student, at the same time point as primary outcome measures are collected (at both pre and post-test).

As trial arm allocation will only be revealed after pre-test outcome measures are collected, the pre-test ASBI scores will be blind to trial arm assignment, despite being collected by classroom teachers. Post-test ASBI scores will not be blind to trial arm assignment. This is unavoidable given the need for a teacher familiar with the student to complete the questionnaire.

The final secondary outcome to be considered as part of this study is the quality of the provision for language and social development, as measured by a composite score based on items from the ECERS-3, ECRES-E and SSTEWS. The items from each scale to be included in this composite measure are detailed in Appendix 1.

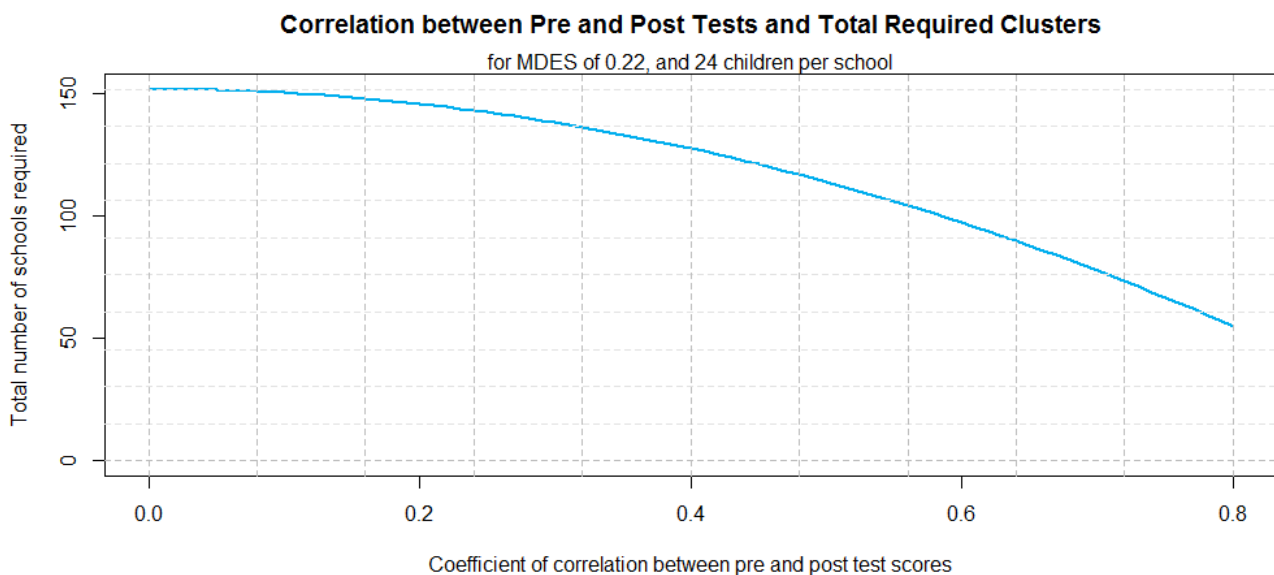
### ***Sample size calculations***

Sample size calculations are based on the assumptions below.

- **Randomisation will be performed at the school-level.** This means that all children in a class will be in the same trial arm, a requirement of this trial given we are testing the effect of teacher training and mentoring, which will impact on whole class attainment.
- **There will be two trial arms (a treatment and control) of approximately equal size.**
- Hypotheses:
  - **Null hypothesis:** There will be no difference in composite language skill scores between children whose teachers experienced ERS training and mentoring and those whose teachers did not.

- **Alternative hypothesis:** There will be a difference in composite language skills scores between children whose teachers experienced ERS training and mentoring and those whose teachers did not (i.e. a two-sided alternative hypothesis).
- **Number of children per cluster is 24.** This is an estimate of the average number of children in each class - the actual number will vary significantly by area, so we have deliberately been conservative.
- **The required minimum detectable effect size (MDES) is 0.22.** This specifies the minimum effect size our trial is powered to detect, in terms of a given standardised difference between two means (of a continuous outcome measure). If the effect of the intervention is below this amount, our trial may not be able to detect it.
- **An intraclass correlation coefficient (ICC) of 0.20.** This defines how alike individual children are within each school (the cluster unit of randomisation). The ICC increases the more individuals within the clusters resemble one another. An ICC of 0.20 is commonly used in clustered randomised control trials in school settings.
- **Power: 80%; Significance level: 5%.** These are standard assumptions.

As we will perform a pre- and post-test of our outcome measures, the predictive power of the pre-test will also factor into our sample size calculations. For the specific language tests used in this trial we do not have information on test-retest correlation over the length of time this study run for. Below we show how the test-retest correlation coefficient impacts the number of schools required for this trial, given the other assumptions listed.





Given test-retest correlation coefficients found in language development studies of similarly aged groups over similar lengths of time<sup>2</sup>, we assumed a test-retest correlation coefficient of 0.50. This requires 114 schools to be enrolled in this trial (no allowance for cluster-level attrition). With 15% attrition at the student-level, (effectively 20 students per cluster, rather than 24), 117 schools would be required.

As our recruitment target is 120 schools, student-level attrition of the degree normally seen in EEF trials is not a threat to our study.

### **FSM power calculations**

Using the assumptions above – including enrolment of 114 schools - we calculate an MDES for FSM students. This assumes that 4 students per cluster are FSM eligible (based on the 16.7% of students receiving FSM as determined by the average across Primary Schools on the EduBase dataset). Given the prioritisation of schools with a larger proportion of FSM pupils, we would hope these calculations are conservative. The assumption of 4 FSM students per cluster gives an MDES of 0.2884.

## **Analysis plan**

Our primary analysis will focus on composite language skill score, and will be performed using Stata. Using the composite score of four language development measures (the APT provides two separate measures) will allow for a more holistic measure of language development, covering comprehension, vocabulary and expressive language. Results for each of the four separate measures will also be reported.

Outcome variables will be regressed using a least squares linear model with treatment arm indicators, strata indicators, and pre-test composite language skill score.. To account for the experimental design, standard errors are clustered at the school level to allow for correlation of pupil outcomes within schools.

We will report the distribution of missing observations by treatment arm and explore whether baseline characteristics are balanced across trial arms.

The estimated impacts will be “intention to treat” (ITT) effects and will be reported with 95% confidence intervals. Effect sizes will be calculated using the Hedges’ g formula. This will require estimates of the standard deviation for the treatment and control groups, which can be derived from the estimated regression. Intra-cluster correlations will be reported.

Secondary analysis of social-behavioural change will use teacher-generated ASBI total scores.

Secondary analysis of the impact of changes in ERS scores on primary and secondary outcome measures will also be reported. The primary measure will be a composite of the items used within the intervention (i.e. those relating to the quality of support for children’s language and social-behavioural development)

---

<sup>2</sup> See EEF (2016) *Nuffield Early Language Intervention: Evaluation report and executive summary*

– this composite ERS measure is described in Appendix 1. Total scores for the ECERS-3 and SSTEW, alongside the literacy subscale of the ECERS-E, will also be reported in our evaluation. Where more than one reception class teacher is participating, BIT will select a class at random.

ERS scores for reception classes will be collected by A+ staff prior to randomisation of schools to trial arms, and again in the autumn term 2017. In 2017 the A+ staff who collect these measures will be coordinated by BIT in order to ensure they are blind to trial arm assignment. For the analysis we will estimate models of the outcome including both pre- and post-intervention ERS scores as covariates; the coefficient on the post-intervention ERS score provides an indication of the extent to which the treatment effect on the outcome has worked through improvements in environment captured by the ERS score, while the residual coefficient on the treatment indicator indicates the extent to which the treatment appears to have improved outcomes through other routes. The impact of treatment on ERS scores will also be analysed. We will also assess the impact of the intervention on pupils who attend a nursery class with an ERS trained teacher, but whose reception year teacher did not receive this training (likely a minority of students in the intervention group).

We will also conduct the analysis for the following subgroups of pupils, using the same model as our primary analysis:

1. Those who are registered for free school meals (FSM) in the National Pupil Database (using the FSMever variable);
2. Those who are marked as English as an Additional Language (EAL) by their schools;
3. Those with language difficulties, as defined as those who score in the bottom 15 percent of BPVS age-standardised scores (in the “extremely” or “moderately” low score range) during pre-test. This is equivalent to a score one standard deviation below the mean of the normed population.
4. The analysis will also be conducted separately for boys and girls.

## Process Evaluation

The process evaluation of the intervention takes a triadic approach through the analysis of the implementation of the intervention; delivery of the intervention and perceived impact of the intervention. Within this structure we seek to determine:

1. Intervention Delivery: To what extent the intervention training and mentoring supports teachers to develop a research led evaluative mind-set and use practical strategies to implement the ERS items within their classroom?
2. Intervention Engagement: To what extent does the teacher engage with the intervention as a tool to catalyse improvement in language and social-behavioural outcomes for children?

A+ Education: Intervention delivery and support: training days; practical resources; mentor support (virtual and physical)

Treatment Schools: Understanding of the intervention; tailoring and adoption of the intervention to the classroom; cascading of the intervention to other members; targeted changes to classes based upon the intervention; improvement of child language and social-behavioural outcomes

It explores this through a range of sub-questions:

### **Intervention Engagement**

- I. To what extent did the participant use, adapt and engage with the intervention, in line with the intervention aims? (Fidelity). See Appendix 2 for definitions of participant engagement.
  - a. What are the barriers to, and conditions of, high and low engagement?
  - b. What barriers prevented schools from achieving the minimum engagement requirements, or surpassing minimum engagement? (Adaptation)
- II. Which ERS composite measures and the pedagogical strategies introduced during the professional development were adopted by schools? What was their rationale? (Dose)
- III. To what extent did the intervention cascade to support staff and other classroom teachers? See Appendix 2 for cascading definitions. (Reach).
  - a. What are the barrier to, and conditions for, cascading the intervention?
- IV. How was the intervention perceived by teachers, senior leaders and teaching assistants?
- V. Did the intervention catalyse change in the school (affective, skills and knowledge, behaviour)?

### **Intervention Delivery**

- VI. How did the mentors support schools to understand, adopt, cascade and embed the intervention?
  - a. How were issues resolved?
- VII. How does the training develop teacher engagement in the intervention?
  - a. How were resources used?
- VIII. How did the training and delivery support change (affect; skills and knowledge and behaviour?)
  - a. Did the intervention catalyse any unexpected changes?
  - b. How do the nursery and reception classes, using the scales to improve language, differ from the control group?

### On Treatment Implementation Minimum<sup>3</sup>

The follow criteria have been outlined as the minimum level of engagement required in order for the intervention to be considered to be taking place. Schools and teachers failing to attain this level of engagement may be considered differently by the process evaluation team. A+ Education will provide web analytic and attendance data. BIT and A+ will work together to gather the ERS items from the mentors. The exact data collection method will be devised in collaboration with the mentors and will be piloted over the autumn term 2016

1) Attendance at training. Teachers must

- a) Attend four of the five training sessions on average, with no teacher attending fewer than three days
- b) Take up the minimum mentor support allocated (three face-to-face visits, at least two of which are in person and mentor observation day)
- c) Attend the follow-up day in the autumn term (at least 50% of teachers)

2) Engagement with intervention. Teachers must:

- a) *Each use at least five ERS items to self-evaluate provision*
- b) *Report or provide concrete evidence (e.g. completed pro-formas, photos) of at least three changes within their classroom as a result of their participation in the intervention*
- c) Access online materials at least once, as must one of their support staff

It is worth noting that the numbers outlined above are currently considered as binary, but when we receive the data we will evolve the binary scale into a sliding scale that accurately reflects the data we receive. For example, accessing online materials once, maybe be too low a measure to accurately capture schools which are not on-treatment.

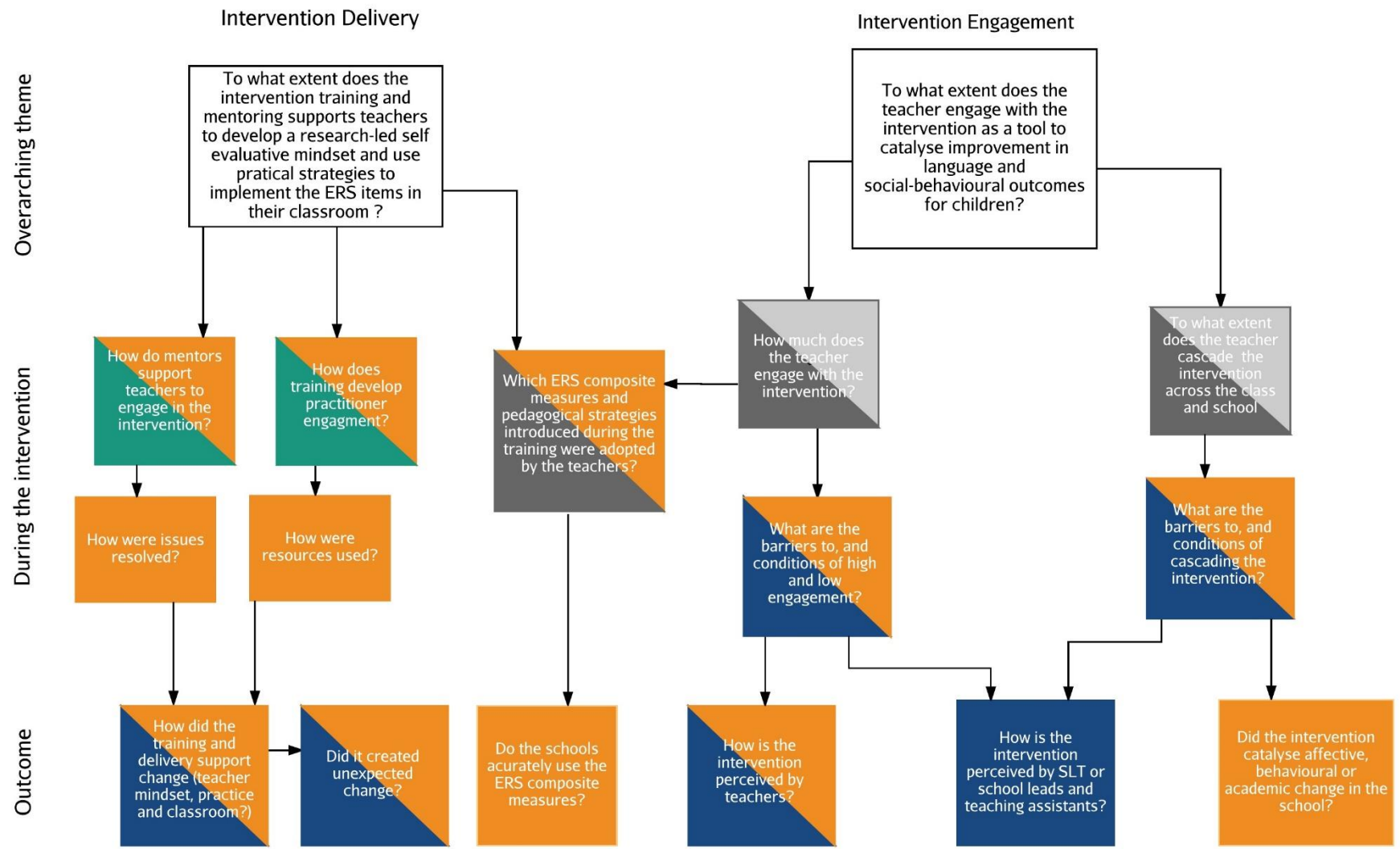
### Methods

We will answer these process evaluation questions through a combination of methods. We will discuss them chronologically as they arise across the project. Please refer to the methods map, which outlines the data collection methods that will be used to answer the specific questions. A chronological timeline is outlined below this.

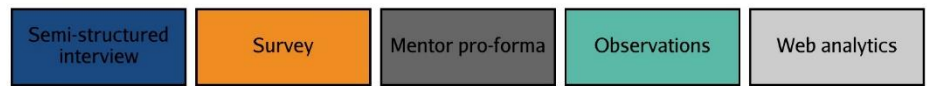
---

<sup>3</sup> This measure assesses the minimum standards required in order for the delivery team to be satisfied that it is on-treatment – it is not an assessment of quality of engagement. The purpose of this measure is to be able to exclude schools which have not engaged in the intervention in the way we expected, so that they can be excluded in the analysis if effect sizes are low as well as contextualise the process evaluation. All will also double up as continuous measures to assess the range of basic engagement within the sample.

# URLEY Implementation and Process Evaluation Methods Map



### Data Collection Method Key



## Chronology of Data Collection

### September 2016

'Business as Usual' questions will be included in the original data ask from schools (pre-randomisation). This will help us determine the differences between schools within both groups, including pre-existing and planned Local Authority support, supported by the delivery team.

### January- June 2017

**Observations** of practitioner training and the practitioner follow up training will help us determine how well the professional development has been delivered, how it improves attendees understanding of the intervention as a tool and provides them clarity around the next steps. This will be triangulated with a documentary analysis of programme materials, training resources, appraisal tools and any other documents that A+ education feel are relevant to share.

### July-November 2017

**Five Case Studies** will be carried out across a range of nursery settings. For this sample, the attributes we aim to use are: ERS scores, Geography; Ofsted; Community Characteristics and Engagement in training to support our purposive sampling. The case studies will be based on triangulated interviews from the classroom teacher who received the training, their mentor and school leader who has strategic responsibility for the intervention. Additional interviews will also be sought with a teaching assistant, other nursery or reception teachers. A minimum of four interviews will be conducted per setting and will use documents from the intervention as prompts, like the reflective folder. Documentary analysis of the reflective evidence folder; ERS scoring (self, mentor and independent); online data and field notes may also be drawn upon to triangulate and contextualise the interview data.

The mentor interviews will add additional insight by providing an overview of all the schools they support. The interviews will be used to ascertain the fidelity, reach, adaption, quality of mentoring and impact measures. All interviews will be digitally recorded with the participants' permission and the recordings transcribed. Data from the observations will be collected on a proforma designed for the evaluation. Individuals will be made aware that their participation is voluntary and of the confidentiality and anonymity of the research.

### June 2018

**A survey** across all schools (both control and treatment), completed by one staff member who has day-to-day involvement with the intervention. To gather data from all participating settings, we propose that an online survey of control and treatment settings be carried out. This would include information on "business as usual" and differences between "business as usual" and the intervention, additional cost data, and gain a wider view of fidelity and/or impact as measured qualitatively. The survey will have a combination of structure and free text questions and will be collected using Survey Monkey. To encourage participation and minimise the burden on respondents it is expected that the survey would take individuals 15 minutes to complete.

**Qualitative analysis:** The data will be analysed using thematic analysis to draw out recurring themes and triangulated across research methods. Additionally, fieldwork notes from baseline and follow up ERS observations will be analysed to understand if particular comments relate to certain positive or negative outcomes. In late 2017 we plan to request self-assessed ERS ratings from participating teaching

staff, so that we may examine how teacher self-assessments correspond to the ratings given by independent observers on the 7 point scale from 'Inadequate to Excellent'..

### **Costs**

An estimate of the per-pupil cost of the intervention will be calculated by the evaluation team. This estimate will focus on cost from the perspective of a participating school and will be based on the marginal, financial costs of the intervention.

The cost estimates will in part make use of information from the project team (particularly with regard to the actual cost of delivering the intervention, e.g. the cost of providing the training), as well as that collected directly by the evaluation team. Information on costs, especially any hidden costs or resource implications, will be explored through the process evaluation as part of the interviews with teachers and school visits. The purpose of collecting such data in the process evaluation would be to identify the main areas of expenditure required by the project. This process will also help to establish whether it may be appropriate to include any questions on costs/resource use in the survey. This will need to strike a balance between collecting sufficient cost information and not damaging response rates; it will also need to take account of whether a teacher is well placed to provide accurate information on particular types of costs.

Costs in terms of time will be reported separately from the financial costs – such as the amount of time for which schools need to arrange supply cover for teachers to attend training. Any costs in terms of prerequisites will also be considered.

An estimate of cost-per pupil per year will also be calculated based on a 3-year time period, as once trained, teachers would also be able to deliver the programme in subsequent years. Any costs associated purely with the trial will be excluded.

## **Ethics and registration**

We take seriously the ethical issues raised in both the quantitative and qualitative elements of the research. NIESR adheres to the Ethics Guidelines of the Social Research Association (SRA). All members of the process evaluation team have Disclosure and Barring Service (DBS) clearance. An ethical review of the project has been undertaken and the project approved by the University of Oxford. While NIESR has an ethics committee consisting of Trustees, we do not consider that this trial requires such additional clearance. The Behavioural Insights Team have an internal ethics process as well as adhering to the ethical standards for research.

The trial protocol is registered on the ISRCTN Trial Registry. The registration number is...

As the intervention will be delivered within school hours, consent from the school should be sufficient with regard to consent for the intervention; and as randomisation is taking place at the school level (rather than randomising individual pupils), the decision to enter into randomisation can also be made by the school. However, we will be collecting personal information on pupils as well as their performance in assessments. We will also be applying for, using and linking to data from the National

Pupil Database. Whether participants are in the treatment or control groups will be identified from the school that they attend. Participants' confidentiality and anonymity will be safeguarded by the methods that we have in place. Given the young age of the children involved in the study, we judge that it is appropriate for the trial to use an opt-in consent process, with participants' parents (or legal guardians) making an informed decision regarding whether they consent to their child's participation in the assessments and data sharing based upon the information provided to them.

All the research assistants used will have enhanced DBS clearance.

A letter sent to parents attached to the opt-in consent form provides information on the aims of the research and the use of data in order that parents are able to make an informed decision about whether to withhold consent from data sharing. The form itself makes it clear and simple for parents to ensure their child(ren)'s data is not shared, and/or that their child does not participate in the assessments, if they do not wish them to do so.

Given the personal information collected in the course of this trial, data security is of utmost importance. As such, it will be transmitted and stored using the security principles underlined in the data security policies of the organisations involved (attached in Appendix A) and the procedures set out in further detail for this specific project in a Data Sharing Agreement. This includes secure transfer of data and use of password-protection and encryption as appropriate during data storage.

## Personnel

### Delivery team

- Sandra Mathers (Oxford)
- Iram Siraj (UCL IoE)
- Clare Williams (A+ Education Ltd)
- Maria Evangelou (Oxford)

### Evaluation team

#### BIT:

- Michael Sanders (Principal Investigator)
- Daniel Carr
- Jessica Heal (Process Evaluation)
- Jake Anders (UCL - consulting to BIT on data analysis)

#### NIESR:

- Anitha George (Process Evaluation design)
- Lucy Stokes
- Richard Dorsett
- Jonathan Buzzeo (Process Evaluation)

### Responsibilities



Outcome measures administration and collection - BIT

Design of the trial

- sample size calculations – BIT and NIESR
- refinement of randomisation approach – BIT and NIESR

Delivery of the intervention

- recruitment of schools – Oxford/UCL/A+ Education
- delivery of training – Oxford/UCL/A+ Education

Data collection

- collection of school and pupil data – BIT and NIESR
- outcome measure collection (research assistant recruitment and coordination) - BIT
- linking of UPN to NPD – NIESR
- data for process evaluation – Oxford/UCL/A+ Education , BIT and NIESR
- data for cost evaluation – Oxford/UCL/A+ Education , BIT and NIESR

Impact analysis – BIT and NIESR

Qualitative analysis – BIT and NIESR (specifically Jessica Heal and Jonathan Buzzeo)

## Risks

The data security policies of Oxford, BIT and NIESR have been provided to the EEF.

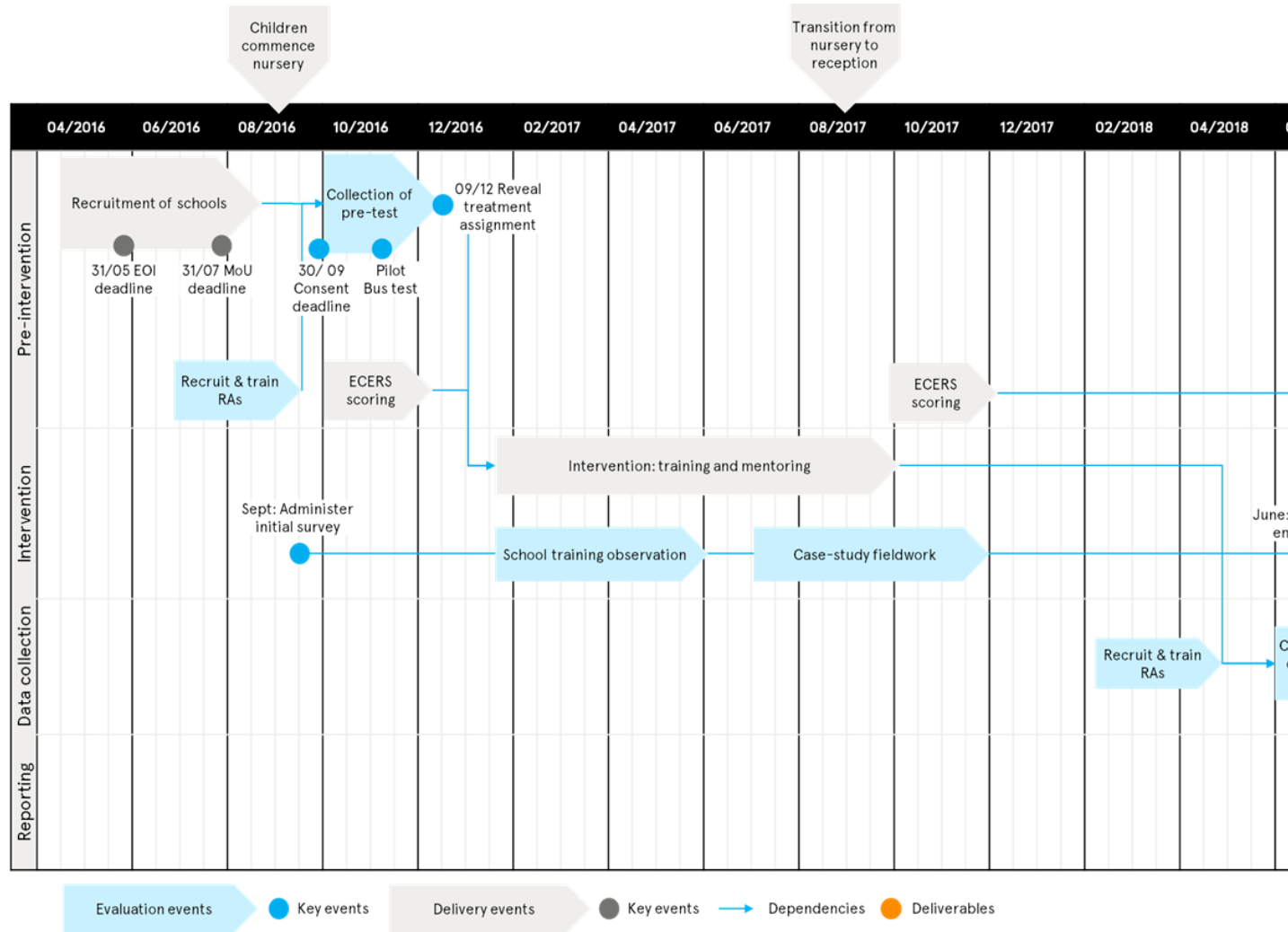
Some of the key risks are listed below:

- School drop-out after randomisation reduces the integrity of the experimental design. To reduce the risk of drop-out, it will be important to ensure schools are well-informed about the programme and the trial from the start, so that they are clear as to what is expected of the evaluation before they commit to taking part. Schools will be asked to sign a memorandum of understanding as a signal of their commitment. It will also be important to maintain good communications with schools throughout the project in order to maximise retention. Drop-out of control schools is a particular risk; to help minimise this control schools will be offered a cash sum, which they will be able to use to purchase the programme at later date, or to spend as they wish.
- There may also be difficulties in recruiting schools to the trial. BIT and NIESR will work closely with the delivery team to convey the importance of the programme to schools and the value to them of them taking part.
- If individuals do not consent to data sharing, or participating in assessments, this has the potential to reduce the sample size, and affect the internal and external validity of the trial. As

consent is collected pre-randomisation, it should not affect the internal validity of the trial, as any withholding of consent should be just as prevalent in the treatment and control groups.

- If pupils are not present on the day of testing this may also reduce the sample size by reducing the number of pupils for whom we are able to obtain a post-test; furthermore, it may introduce some bias if it is a non-random group of pupils who are absent. However, the recruitment target of 120 schools provides some allowance for pupil-level attrition.
- There is a possibility that the delivery of the intervention will vary across schools. However, this reflects the reality of implementing such a programme; impact estimates therefore relate more to type of treatment likely to prevail in practice rather than that which might be observed under ideal conditions. Nevertheless understanding treatment variation is important and will be explored as part of the process evaluation.
- When randomising clusters rather than individuals the chances of a 'bad draw' (e.g. proportionally more high FSM share schools in one trial arm) increase because of the smaller number of units. We will use blocking to limit this problem.

# Project schedule



## References

- Anderson, L. M., Shinn, C., Fullilove, M. T., Scrimshaw, S. C., Fielding, J. E., Normand, J., Carande-Kulis, V.G & Task Force on Community Preventive Services. (2003). The effectiveness of early childhood development programs: A systematic review. *American journal of preventive medicine*, 24(3), 32-46.
- Bierman, K. L., Domitrovich, C. E., Nix, R. L., Gest, S. D., Welsh, J. A., Greenberg, M. T., ... & Gill, S. (2008). Promoting academic and social-emotional school readiness: The Head Start REDI program. *Child development*, 79(6), 1802-1817.
- Bierman, K. L., Nix, R. L., Heinrichs, B. S., Domitrovich, C. E., Gest, S. D., Welsh, J. A., & Gill, S. (2014). Effects of Head Start REDI on children's outcomes 1 year later in different kindergarten contexts. *Child Development*, 85(1), 140-159
- Campbell, F., Conti, G., Heckman, J. J., Moon, S. H., Pinto, R., Pungello, E., & Pan, Y. (2014). Early childhood investments substantially boost adult health. *Science*, 343(6178), 1478-1485.
- Currie, J. (2001). Early childhood education programs. *The Journal of Economic Perspectives*, 15(2), 213-238.
- Domitrovich, C. E., Cortes, R. C., & Greenberg, M. T. (2007). Improving young children's social and emotional competence: A randomized trial of the preschool "PATHS" curriculum. *The Journal of primary prevention*, 28(2), 67-91.
- Gorey, K. M. (2001). Early childhood education: A meta-analytic affirmation of the short-and long-term benefits of educational opportunity. *School Psychology Quarterly*, 16(1), 9.
- Heckman, J.J., (2012). The case for investing in disadvantaged young children. European Expert Network on Economics of Education.
- Heckman, J.J. and Masterov, D.V. (2007). The productivity argument for investing in young children. *Applied Economic Perspectives and Policy*, 29(3), pp.446-493.
- Hopkin R, Stokes L and Wilkinson D (2010) Quality, Outcomes and Costs in Early Years Education, Report to the Office for National Statistics
- Raver, C. C., Jones, S. M., Li-Grining, C., Zhai, F., Metzger, M. W., & Solomon, B. (2009). Targeting children's behavior problems in preschool classrooms: A cluster-randomized controlled trial. *Journal of Consulting and Clinical Psychology*, 77(2), 302.
- Sibieta, L., Kotecha, M. & Skipp, A. (2016) Nuffield Early Language Intervention. Evaluation report and executive summary, Education Endowment Foundation.
- Springate, I., Atkinson, M., Straw, S., Lamont, E. & Grayson, H. (2008) *Narrowing the gap in outcomes: early years (0-5 years)*, Slough: NFER.
- Sylva, K., Melhuish, E.C., Sammons, P. Siraj, I. and Taggart, B. (2008). Final Report from the Primary Phase: Pre-school, School and Family Influences on Children's Development during Key Stage 2 (7-11). Nottingham: DCSF Research Report 61 / Institute of Education, University of London
- Sylva, K., Melhuish, E.C., Sammons, P., Siraj, I. and Taggart, B. with Smees, R., Toth, K. & Welcomme W. (2014) Effective Pre-school, Primary and Secondary Education 3-16 Project (EPPSE 3-16) Students' educational and developmental outcomes at age 16, Department for Education Research Report RR354

## Appendix 1: ERS composite measure items

Items have been chosen on the basis that they reflect the behaviours/practices expected to be most closely related to the target outcomes:

- Vocabulary (BPVS)
- Comprehension (CELF sentence structure)
- Expression (Renfrew APT)
- Social behavioural development (ASBI)

There are other items within the ERS which may be used in the professional development. For example, STEW Item 11 (Encouraging sustained shared thinking and exploration) is likely to be used as part of a focus on scientific activities, which provide a rich context for language. This has not been included as a core item, as it provides a *vehicle* for improving opportunities for communication and language rather than representing target behaviours in its own right. Any resulting increase in support for communication and language should be picked up through change in the core items.

The ECERS-3 and -E items relating to use of books are very similar. A mapping exercise has been carried out to inform the choice of items, resulting in the choice of ECERS-3 Item 14 (but not 15) and ECERS-E Item 2 (but not 3). The mapping can be provided if required.

### Summary of core 18 items to be included in the 'URLEY' composite measure

<p><b>ECERS-3 (8 items)</b></p> <ul style="list-style-type: none"> <li>• 20 Blocks</li> <li>• 21 Dramatic play</li> <li>• 22 Nature/science</li> <li>• 30 Staff-child interaction</li> <li>• 31 Peer interaction</li> <li>• 12 Helping children expand vocabulary</li> <li>• 13 Encouraging children to use language</li> <li>• 14 Staff use of books with children</li> </ul> <p><b>ECERS-E (2 items)</b></p> <ul style="list-style-type: none"> <li>• 2 Book and literacy areas</li> <li>• 6 Talking and listening</li> </ul>	<p><b>SSTEW (8 items)</b></p> <ul style="list-style-type: none"> <li>• 1 Self-regulation and social development</li> <li>• 4 Supporting socio-emotional wellbeing</li> <li>• 5 Encouraging children to talk with others</li> <li>• 6 Staff actively listen to children and encourage children to listen</li> <li>• 7 Staff support children's language use</li> <li>• 10 Encouraging SST through storytelling, sharing books, singing and rhymes</li> <li>• 8 Sensitive responsiveness</li> <li>• 14 Assessing language development</li> </ul>
---	--

### Full list of 'URLEY' composite measure

Bold = item to be included in composite

### ECERS-3 (8 items)

- Learning activities
  - (17 Fine motor)
  - (18 Art)
  - (19 Music and movement)
  - **20 Blocks**
  - **21 Dramatic play**
  - **22 Nature/science**
  - (23 Maths materials and activities)
  - (24 Maths in daily events)
  - (25 Understanding written numbers)
  - (26 Promoting acceptance of diversity)
  - (27 Appropriate use of technology)
- Interactions
  - (28. Supervision of gross motor)
  - (29 Individualised teaching and learning)
  - **30 Staff-child interaction**
  - **31 Peer interaction**
  - (32 Discipline)
- Language and literacy subscale
  - **12 Helping children expand vocabulary**
  - **13 Encouraging children to use language**
  - **14 Staff use of books with children**
  - (15 Encouraging children's use of books)
  - (16 Becoming familiar with print)

### ECERS-E (2 items)

- Literacy subscale
  - (1 Environmental print)
  - **2 Book and literacy areas**
  - (3 Adult reading with children)
  - (4 Sounds in words)
  - (5 Emergent writing/ mark making)
  - **6 Talking and listening**

### SSTEW (8 items)

- Building trust, confidence and independence
  - **1 Self-regulation and social development**
  - (2 Encouraging choices and independent play)
  - (3 Planning for small group and individual interactions/adult deployment)

- Social and emotional wellbeing
  - **4 Supporting socio-emotional wellbeing**
- Supporting and extending language and communication subscale
  - **5 Encouraging children to talk with others**
  - **6 Staff actively listen to children and encourage children to listen**
  - **7 Staff support children's language use**
  - **8 Sensitive responsiveness**
- Supporting learning and critical thinking subscale
  - *(9 Supporting curiosity and problem solving)*
  - **10 Encouraging SST through storytelling, sharing books, singing and rhymes**
  - *(11 Encouraging SST in investigation and exploration)*
  - *(12 Supporting children's concept development and higher-order thinking)*
- Assessing learning and language
  - *(13 Using assessment to support and extend learning and critical thinking)*
  - **14 Assessing language development**

## Appendix 2: Success Criteria to Assess Fidelity

This follow is examples provided by the delivery team which can be used as a steer to help assess fidelity.

### Delivery of Intervention

- Delivery team deliver the specified and manualised aspects of the intervention with fidelity and show evidence of a system for adapting the flexible aspects of the intervention according to an agreed framework

### 3 Stages of Participant Engagement

1. Participating teachers are aware of and can articulate the principles of the intervention in a way which demonstrates understanding
2. Participating teachers have used the ERS items to self-evaluate provision and help them decide on potential areas for development
3. Participating teachers have implemented the evidence-based strategies presented to them within the professional development, in order to address areas identified through their use of the ERS as being in need of development
  - Participating teachers have made structural changes within their classrooms (e.g. room layout, resourcing, organisation of the day)
  - Participating teachers have made process changes within their classrooms (e.g. pedagogical strategies, changes to the ways in which staff interact with the children and support their learning)

### Cascading the Intervention

1. Participating teachers have made efforts to cascade the intervention to other members of staff (relevant variables include number of staff engaged, number of efforts to engage and involve other staff, and quality of efforts)
2. Teaching assistants and other staff within participating teachers' classrooms are aware of and can articulate the principles of the intervention in a way which demonstrates understanding, have been engaged in the change process
3. Teachers and other staff within their classrooms access the online materials and can describe the ways in which they have used them
  - Participating teachers and other staff within their classrooms show evidence of having adapted their implementation on the basis of interactions with their mentor