Concept Cat: A two-armed cluster randomised controlled trial Evaluation Protocol

Evaluator (institution): RAND Europe, University of York
Principal investigator(s): Elena Rosa Brown, Miguel Subosa, Louise Tracey, Erin Dysart

Education Endowment Foundation

Evaluation summary

Project title	Concept Cat: A two-armed cluster randomised controlled trial
Developer *(Institution)*	Better Communication CIC
Evaluator *(Institution)*	RAND Europe, University of York
Principal investigator(s)	Elena Rosa Brown, Miguel Subosa, Louise Tracey, Erin Dysart
Protocol author(s)	Fin Oades, Miguel Subosa, Elena Rosa Brown, Erin Dysart, Louise Tracey
Trial design	Two-arm cluster randomised controlled trial with random allocation at setting level
Trial type	Efficacy
Pupil age range and Key stage	3-4, Early Years Foundation Stage 1
Number of settings *(at design stage)*	90
Number of pupils *(at design stage)*	A minimum of 15 children per setting, or 1350 children in total
Primary outcome measure and source	*Basic Concepts* sub-test score from the Clinical Evaluation of Language Fundamentals® Preschool-2 (CELF Preschool-2)
Secondary outcome measure and source	Early Years Toolbox Early Numeracy task *Concepts and Following Directions* sub-test score from the CELF Preschool-2

Protocol version history

Version	Date	Reason for revision
1.0 [*original*]	07/09/2023	N/A

Contents

Evaluation summary 1

Study rationale and background 4

Intervention 5

Impact evaluation design 9

- Research questions 9
- Design 10
- Participant selection 11
- Outcome measures 12
- Sample size 15
- Randomisation 17
- Statistical analysis 18

Implementation and process evaluation (IPE) design 21

- Research questions 21
- Research methods 22
- Analysis 26

Cost evaluation design 36

Ethics and registration 36

Data protection 37

Personnel 38

Risks 40

Timeline 42

References 43

Appendix A: Wave 1 (pre-trial IPE timeline) 45

Appendix B: Wave 2 (main IPE timeline) 46

Appendix C: Identification of children with higher needs within Concept Cat programme 49

Appendix D: Data flow diagram 53

Appendix F: Information sheet and withdrawal form 61

Appendix G: Logic model mapped to implementation dimensions and research methods 68

Tables

Table 1: Trial design 10

Table 2: Sample size calculations 16

Table 3: Research questions mapped to implementation and research methods 27

Table 4: IPE methods and analysis overview 29

Table 5: Timeline 42

Figures

Figure 1: Theory of Change for Concept Cat, developed with facilitation from EEF ToC workshop 6

Study rationale and background

There is evidence to suggest a link between conceptual language development and maths skills. Specifically, children who are exposed to more conceptual language in their early years (EY) tend to have stronger maths skills later in life. One study found that children who were exposed to more maths-related language in their homes and preschools had better maths skills in kindergarten than children who were exposed to less maths-related language (LeFevre et al., 2010). Another study found that children who were exposed to more spatial language (which is a type of conceptual language) in their preschool years had better spatial reasoning skills in elementary school (Verdine et al., 2014).

Evidence also suggests that children from disadvantaged backgrounds tend to have lower levels of conceptual vocabulary than their more advantaged peers, which in turn can have negative effects on their language, cognitive, and academic development. For example, a landmark study by Hart and Risley (1995) found that children from low-income families heard significantly fewer words overall than children from higher-income families, and that the words they did hear tended to be more limited in scope and complexity. This disparity in language exposure was found to have long-lasting effects on children's language and cognitive development, with children from low-income families having smaller vocabularies and weaker language skills overall.

Other studies have found similar patterns of language disparities between children from different socioeconomic backgrounds. For example, a study by Fernald et al. (2013) found that children from low-income families had lower levels of exposure to both conversational and conceptual language than children from higher-income families, and that this disparity was evident as early as 18 months of age.

In comparison, existing research strongly suggests that early language interventions have the potential to a major impact on shaping language skills (Law, et al., 2017; Fricke et al., 2012; Scarborough, 2009; Marulis & Neuman, 2012) with a recent metanalysis of EY language and communication-focused programmes demonstrating potential for high positive impact on children (EEF 2023b). Therefore, there is a compelling case to develop and evaluate interventions that support conceptual language development, particularly for disadvantaged children.

In the UK, however, few programmes have been evaluated with sufficiently robust methodologies. Thus, robust evaluations of EY language development programmes in the UK would be an invaluable addition to the evidence base.

This evaluation also forms part of the Department of Education's (DfE) Stronger Practice Hubs (SPH) policy which are designed to build evidence-informed practice in EY. The SPHs, launched in November 2022, form part of the DfE's Early Years Education COVID-19 Recovery Package. A key aim of SPHs is to address the impact of the pandemic on young children by supporting EY settings to build local networks and share evidence-informed practices to ultimately improve the quality of education and care. The EEF is supporting the launch of the SPHs to build evidence around EY approaches. EEF's work includes:

1. Selecting programmes from its open funding rounds to be part of the list of programmes that Stronger Practice Hubs are able to make available as funded support in their region.
2. Providing funding, in addition to funding from Stronger Practice Hubs, for programme providers to deliver their programme to EY settings as part of a research project.
3. Providing funding for independent evaluators to implement research of programmes on the list available to Stronger Practice Hubs

This evaluation will help to build evidence around what works in EY and will be run in three stronger practice hubs in Birmingham, Trafford, and Everton.

The planned efficacy trial for Concept Cat will be carried out as a two-arm, waitlisted, cluster randomised controlled trial, with a 50:50 allocation of 90 EY settings to treatment and control groups. Baseline testing will take place from September to October 2023, with randomisation in September 2023. The intervention itself will then be delivered from September 2023 until June 2024, corresponding to a duration of 30 weeks.

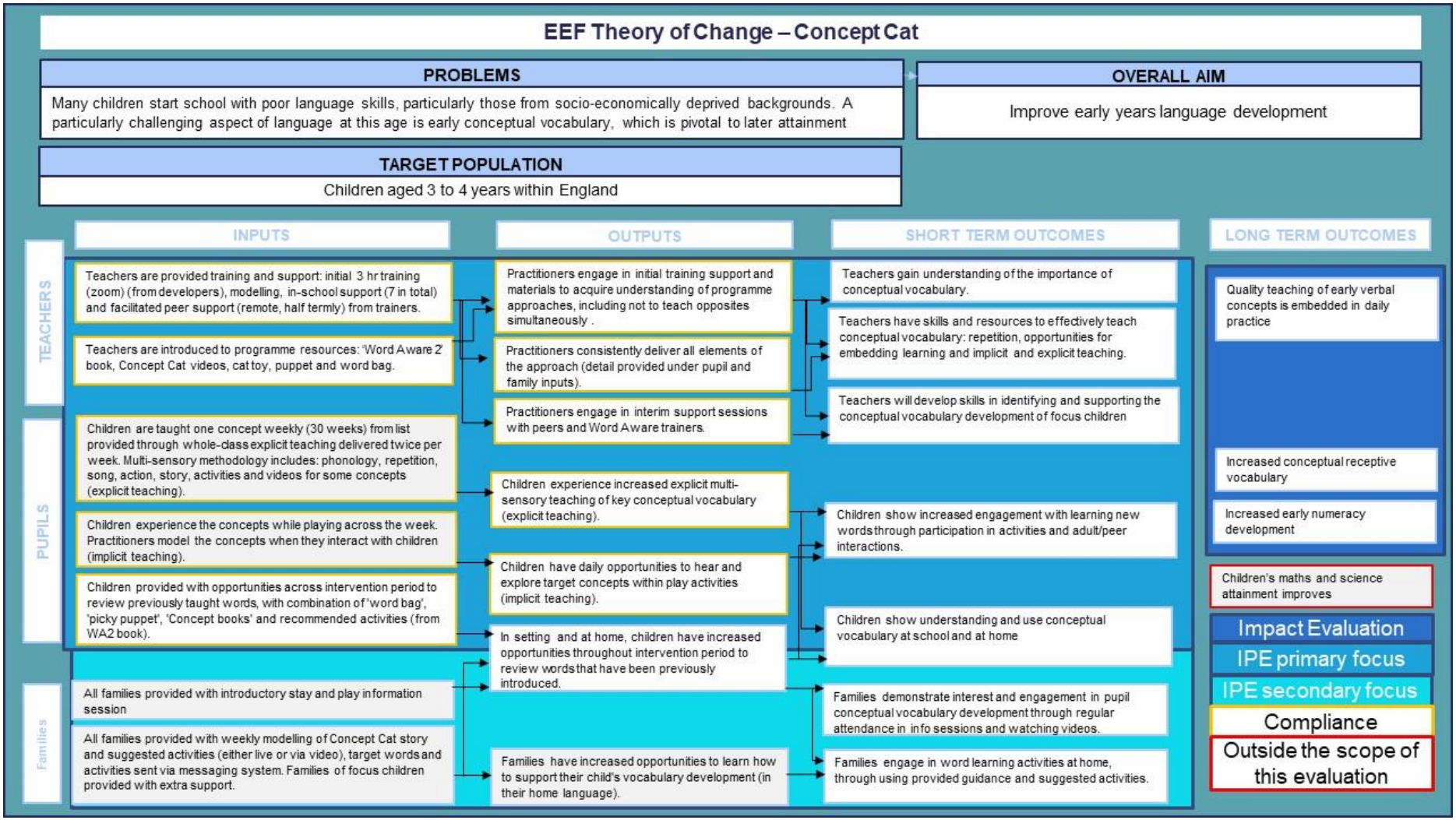
Intervention

Concept Cat was conceived by the founders of Thinking Talking, Stephen Parsons and Anna Branagan, and its teaching methodology is outlined in their book, *Word Aware 2*. The programme is grounded in the STAR (Select, Teach, Activate, Review) approach to EY conceptual learning, with its selection of 'key concepts' being based upon the three-tiered classification of core early vocabulary. Moreover, it seeks to interweave classroom practice with child-parent interactions, an area for which there is only a limited number of relevant previous evaluations.

Concept Cat is already relatively widely delivered, with approximately 300 EY practitioners trained in the Concept Cat teaching methodology each year. A quasi-experimental impact evaluation of the Concept Cat programme showed promise with effect sizes of 0.42. However, there were some distinct limitations in the study, namely the lack of randomisation in the design and the limited sample size, with data having been collected from only two preschools. Furthermore, the programme has been adapted since the pilot, with staff in settings being trained to deliver the programme in this evaluation, compared to the use of speech and language therapy (SALT) students to deliver the programme in the pilot (Hopkins, et al., 2022). These design and implementation choices may have inflated the impact seen in the pilot. As such, there is a need for a methodologically robust efficacy trial implemented at a larger scale to ascertain the true impact of Concept Cat on children.

Concept Cat is a whole-class intervention, targeting children aged between three and four, that seeks to facilitate the acquisition of key early verbal concepts. In turn, the acquisition of these early verbal concepts would support the attainment of learning competencies laid out in the Key Stage 1 core science and mathematics curricula. The Concept Cat methodology offers an alternative to the generally unstructured and less explicit way these core concepts are taught in standard practice. Moreover, the Concept Cat approach offers a combination of explicit and implicit teaching of concept words which is embedded in daily practice. The programme is delivered over a full academic year (i.e., approximately 30 weeks), where each week coincides with the introduction of a new key word. Further details of the programme can be seen in the Theory of Change (Figure 1) below.

***Figure 1**: Theory of Change for Concept Cat, developed with facilitation from EEF ToC workshop*



The programme incorporates four core components in its implementation. These are a whole-class introduction to the word, meaningful play sessions, parent-child tasks, and a whole class review.

The '**whole-class introduction'** establishes the key verbal concept for a given week, forming the 'explicit teaching' aspect of the programme, delivered through a multi-sensory methodology. The specific word is introduced by the teacher and is accompanied by both a unique visual symbol and a physical gesture. Phonology and repetition are also used, supporting the children's memory of the word's phonic characteristics.

'**Meaningful play sessions**' provide an opportunity for more 'implicit teaching' to reinforce the learning that has occurred during the whole-class introductory phase, whereby the children come across the word within structured play situations. Children's families are also introduced to the word, and '**home-based tasks**' are suggested, such as discussing the meaning and use of the word with the child whilst walking home from school. There is also a '**whole-class review**', which incorporates a range of activities using Concept Cat props such as the word bag and picky puppet, which will allow children to encounter the word during both the word's focus week, as well as subsequent weeks, concretising the knowledge the children have gained throughout the intervention's entire delivery period.

Children with the most limited language proficiency are typically identified by comparing their capabilities against the 'Early Learning Goals'; they are provided with additional support through extra modelling of the Concept Cat story and target words. For the purposes of this evaluation, children requiring additional support are referred to as 'focus children'. The following criteria must be satisfied to qualify as a focus child:

- Aged 3-4 years old
- Able to sit and respond to an adult-led task for a few moments
- Uses fewer words and shorter sentences than other children of the same age
- Not a reluctant speaker at nursery who speaks fluently at home
- If the child speaks English as an Additional Language (EAL) then they must also have delayed language development in the home language(s)

The overall structure of the intervention is based on the STAR (Select, Teach, Activate, Review) methodology, where a concept is 'selected', 'taught', then implicitly 'activated' through play and home-based activities, and subsequently 'reviewed' to encourage a deeper understanding of its meaning.

Delivery personnel and training

The programme is delivered by teachers within the selected EY settings, with lead practitioners receiving a three-hour remote training session and other EY staff receiving a one-hour remote training session provided by Better Communication. Each setting will also receive seven in-setting support visits from Concept Cat Coaches. During the initial day visit, setting staff under coach supervision carry out an initial assessment of six children, selected by practitioners (based on a range of language abilities) so the correct level of concept words is selected for the setting. Coaches also model the Teach element of the programme. Other sessions involve further modelling of Teach, Activate, and Review and give further support on implementation of the programme. The sessions provide an opportunity for

coaches to ensure practitioners are implementing the programme as it is intended and in a way that works for the setting. In addition to the coach support sessions, lead practitioners are also encouraged to attend six group support sessions across the academic year, allowing practitioners to share experiences and tips to improve delivery.

Duration and frequency

The intervention will commence in the final week of September, running for a total of 30 active weeks. As discussed above, the intervention is delivered using a weekly (five-day or three-day) structure, where a new word is introduced at the start of each of week for the programme's duration. The explicit teaching aspect occurs twice a week for approximately ten minutes, whilst the other implicit methodologies may be interwoven into other, less prescriptive activities throughout the week.

Materials

The key materials required for delivery are:

- The Word Aware 2 book (Parsons and Branagan, 2016) provides the overall teaching structure for each of the words;
- Concept Cat soft toy, word bag and 'picky' puppet are used for introducing and reviewing the taught words;
- Lift Lessons animated videos of approximately 50% of the introductory stories, and
- in addition, each setting is provided with the printed materials for the word of the week for all of Level 1 and Level 2 words with the option to print locally any Level 3 words needed based on the Concept Cat Screen and as identified by the Concept Cat Coach for each setting.

EY settings are required to find a few resources to incorporate into the introductory Concept Cat stories. These are simple everyday items such as socks or boxes or toys that settings are highly likely to have.

Impact evaluation design

Research questions

The primary research question of this project is:

> **RQ1.** What is the difference in early conceptual vocabulary development, measured by the Basic Concepts subtest of the Clinical Evaluation of Language Fundamentals Preschool-2 UK (CELF-Preschool-2), of pupils in settings receiving Concept Cat in comparison to those pupils in control settings receiving business as usual?

The following sub-questions of this primary research question (RQ1) will also be explored:

> **RQ1a.** What is the impact of the Concept Cat teaching methodology on the early conceptual vocabulary development of Early Years Pupil Premium/Free Early Education Entitlement (EYPP/FEEE)-eligible pupils, compared to non-EYPP/FEEE-eligible pupils?
>
> **RQ1b.** What is the impact of the Concept Cat teaching methodology on the early conceptual vocabulary development of pupils with English as an Additional Language (EAL), compared to non-EAL pupils?
>
> **RQ1c.** What is the impact of the Concept Cat teaching methodology on the early conceptual vocabulary development of pupils with Special Educational Needs or Disability (SEND), compared to non-SEND pupils?

The secondary research questions of this project are:

> **RQ2.** What is the difference in early conceptual vocabulary development, measured by the Concepts and Following Directions subtest of the CELF-Preschool 2, of pupils in settings receiving Concept Cat in comparison to those pupils in control settings receiving business-as-usual?
>
> **RQ3.** What is the difference in early numeracy development measured by the Early Numeracy Assessment (ENA) of the Early Years Toolbox (EYT) of pupils in settings receiving Concept Cat in comparison to those pupils in control settings receiving business-as-usual?

The following sub-questions of this secondary research question (RQ3) will also be explored:

> **RQ3a.** What is the impact of the Concept Cat teaching methodology on the early numeracy development of EYPP/FEEE-eligible pupils, compared to non-EYPP/FEEE-eligible pupils?
>
> **RQ3b.** What is the impact of the Concept Cat teaching methodology on the early numeracy development of EAL pupils, compared to non-EAL pupils?
>
> **RQ3c.** What is the impact of the Concept Cat teaching methodology on the early numeracy development of SEND pupils, compared to non-SEND pupils?

Design

Table 1: Trial design

Trial design, including number of arms		Two-group, waitlisted, cluster randomised controlled trial
Unit of randomisation		Early years settings
Stratification variables		Setting type (Private, voluntary, or independent [PVI] vs. school-based settings [SBS]); Region will also be included as a variable (four in total)
Primary outcome	**Variable**	Early Conceptual Vocabulary
	Measure (instrument, scale, source)	'Basic Concepts' subtest from Clinical Evaluation of Language Fundamentals Preschool-2 UK (CELF-Preschool 2 UK).
Secondary outcome(s)	**Variable(s)**	Early Conceptual Vocabulary, Early Numeracy
	Measure(s) (instrument, scale, source)	Early conceptual vocabulary measured using the 'Concepts and Following Directions' subtest from Clinical Evaluation of Language Fundamentals Preschool-2 UK (CELF-Preschool 2 UK). Early numeracy measured using Early Years Toolbox (EYT) Early Numeracy task.
Baseline for primary outcome	**Variable**	Early Conceptual Vocabulary
	Measure (instrument, scale, source)	'Basic Concepts' subtest from Clinical Evaluation of Language Fundamentals Preschool-2 UK (CELF-Preschool 2 UK).
Baseline for secondary outcome	**Variable**	Early Conceptual Vocabulary 'Basic Concepts' subtest from Clinical Evaluation of Language Fundamentals Preschool-2 UK (CELF-Preschool 2 UK).
	Measure (instrument, scale, source)	Early Conceptual Vocabulary 'Basic Concepts' subtest from Clinical Evaluation of Language Fundamentals Preschool-2 UK (CELF-Preschool 2 UK).

As outlined in Table 1, this evaluation is designed as a two-arm, waitlisted, cluster randomised controlled efficacy trial, with randomisation at setting level and settings stratified according to setting type (PVI vs. SBS) and region. Randomisation will be performed at the level of settings to avoid contamination, as well as to account for Concept Cat's whole-class methodology. Settings randomly allocated to the treatment condition will deliver Concept Cat whilst those randomly allocated to the control condition will proceed with business as usual.

The primary outcome will be the same as that used in the pilot study (early conceptual vocabulary, measured using the Basic Concepts subtest of CELF-Preschool 2; Hopkins et al., 2022). In addition, early conceptual vocabulary will be measured using the *Concepts and Following Directions* subtest from CELF Preschool-2 (secondary outcome) to evaluate the

impact of Concept Cat on children's ability to: (a) understand spoken directions containing concepts that require logical operations; (b) remember names, order, and characteristics of items mentioned; and (c) identify the target from among several choices. Early numeracy will also be measured using the Early Numeracy Assessment (ENA) from the Early Years Toolbox (EYT) to understand how Concept Cat may impact on early numeracy development. More details on outcome measures are discussed in the *Outcome Measures* section.

The evaluation has two phases:

- Wave 1 is the pre-delivery phase, which will take place during Spring and Summer 2023, and will include observations of the three- and one hour-training sessions, four visits to settings with Concept Cat Coaches, interviews with Concept Cat Coaches and focus groups with practitioners (see IPE section).
- Wave 2 is the main trial.

Participant selection

Settings

Settings will be recruited into the trial with support from the delivery team, Better Communication CIC. Recruitment will focus on settings from three Stronger Practice Hubs (SPHs), as follows:[1]

- HEART – Midlands Early Years SPH based in Birmingham in the West Midlands
- Bright Futures North West Early Years SPH based in Trafford in the North West
- An SPH based in Everton in the North West, the name of which is, at the time of writing, still to be confirmed.

Settings are required to have a minimum of 15 children aged 3 to 4 (in Foundation 1) enrolled to attend for at least 15 hours a week in the academic year 2023/2024 and would ideally have a higher-than-average number of EYPP-eligible pupils. However, we anticipate that a number of settings will, exceptionally, have fewer than 15 eligible children, but the variability in the number of eligible children across all settings will even this out.

In addition, the following exclusion criteria will be observed:

- settings where staff have attended Word Aware Early Years training within the last 3 years;
- settings that have implemented Concept Cat within the last 2 years; and
- settings currently accessing Lift Lessons Concept Cat videos.
- However, settings that have the *Word Aware 2* book but are not actively using it are still eligible to participate.

[1] To support with recruitment for the trial, it was felt that a larger recruitment area than that of two hubs would be beneficial to maximise take up. The hub appointment process in Autumn 2022 only resulted in one hub being selected for North West. A second round is being undertaken over the Spring term where the application will specify which programmes have been matched to that area and will be funded by the hub.

Pupils

To ensure that we are evaluating the programme's impact on pupils for whom Concept Cat is believed to be most suitable, the following inclusion will be applied in pupil selection:

- Pupils who are in Foundation 1 and aged three to four during the academic year 2023/2024.
- Pupils who are registered to attend the setting for a minimum of 15 hours per week.

On the other hand, pupils will not be eligible to participate in the evaluation if at baseline:

- they are not between three- to four-years old;
- they are judged by setting staff as unable to sit and follow a short adult-selected task. This is because these needs would prevent them from accessing the assessments;
- they attend the setting less than 15 hours per week; and
- they are unable or unwilling to take part in baseline testing (i.e., because of speech and language difficulties or special educational needs).

For the purposes of the trial, children will be identified in September 2023 (i.e., at the start of the school year). Teachers will be encouraged to use their own judgement and may base this on their own evaluation of how children perform in class activities. Parents and guardians may opt to withdraw their child from the trial at any point without penalty. Upon notification of the parent or guardian's request to withdraw their child, the delivery and evaluation teams will delete the respective child's data and produce a corresponding destruction log. Withdrawals will be communicated across RAND, Elklan, and Better Communication CIC and documented in a shared tracking form that will be accessible to all parties.

A detailed description of the data collection process, as well as corresponding data safeguarding procedures, is included in the 'Data protection' section of this protocol.

Outcome measures

Baseline measures

Baseline testing will be conducted to improve the precision of the impact estimate and explore (and potentially control for) imbalance at baseline. To maximise the correlation between baseline and endline, we will use the same measure at both time points: the *Basic Concepts* subtest of CELF Preschool-2 (see *Primary outcome* measure for more detail).

To minimise burden on pupils and allow sufficient time for intervention delivery, we will not measure early numeracy scores – our secondary outcome – at baseline. Instead, we will use the CELF Preschool-2 as a baseline measure for early numeracy, appreciating that, while pre- and post-test correlations may be low, overlap between the concepts is likely.

Primary outcome

The primary outcome for the efficacy trial will be pupils' early conceptual language development, as assessed by the *Basic Concepts* subtest of CELF Preschool-2. The CELF Preschool-2 is a standardised, individually administered assessment of expressive and

receptive linguistic ability specifically designed for children aged three to six. It is widely used in EY outcome assessments. The CELF Preschool-2 consists of seven subtests, including *Basic Concepts*.[2]

The CELF Preschool-2 was judged fit for purpose since: (a) it is designed to be brief (taking around five to seven minutes to administer per subtest); (b) it directly measures receptive vocabulary development; (c) it is UK norm-referenced; (d) and it has strong psychometric properties, with test-retest reliability ranging from 0.77 to 0.96 (for ages 3 to 3,11) and 0.74−0.95 (for ages 4 to 4:11; EEF, 2023a).

We selected the *Basic Concepts* subtest as our primary outcome since it was also used in the pilot study, thus allowing for comparison across evaluations. We note that the pilot found slight ceiling effects in the Basic Concepts subtest at baseline and understand that this may be attributable to design factors, such as the sampling strategy used (e.g., a slightly older pupil group than the efficacy trial is targeting, lower-than-average number of pupils with English as an Additional Language). While we therefore do not anticipate such effects to persist in the efficacy trial, we propose to use Phase 1 to pilot CELF-P2 UK in settings outside the evaluation to ascertain their appropriateness, explore feasibility of administration, and review psychometric properties before the trial begins.

The *Basic Concepts* subtest evaluates a child's knowledge of the following concepts: dimension and size; direction, location, and position; number and quantity; and quantitative equality. The assessment is performed by providing a description of a concept to the child and asking them to select from a set of options the picture that best corresponds with or exemplifies the said concept. Scores are calculated as the number of correctly identified concepts, and the test is discontinued after five consecutive incorrect responses.

Testing will be administered by Elklan's independent test administrators, who are qualified SALTs. Data will be collected using the paper-based record sheets as per standard delivery of CELF Preschool-2, with data being uploaded by test administrators to a secure portal, hence allowing for ongoing quality assurance by Elklan and the evaluation team. Test administrators will be trained in the use and administration of CELF Preschool-2, including how to conduct practice sessions as well as upload assessment data, and will be blinded to allocation status.

Secondary outcomes

We will further explore the impact of the intervention on two main secondary outcomes:

1. **early conceptual vocabulary** measured using the 'Concepts and Following Directions' subtest from CELF-Preschool 2 UK; and
2. **early numeracy** measured using the EYT ENA.

These outcomes are discussed below in greater detail.

[2] https://www.pearsonclinical.co.uk/store/ukassessments/en/Store/Professional-Assessments/Developmental-Early-Childhood/CELF-Preschool-2-UK/p/P100009267.html

Early conceptual vocabulary

Beyond the development of conceptual language captured by the primary outcome, early conceptual vocabulary will also be measured using children's scores on the CELF Preschool-2's *Concepts and Following Directions* subtest.

The Concepts and Following Directions subtest evaluates a child's ability to: (a) understand spoken directions containing concepts that require logical operations; (b) remember names, orders and characteristics of items mentioned; and (c) identify the target from among several choices. It is distinct from the primary outcome in that it measures conceptual *vocabulary* instead of directly measuring concepts and is linked to receptive vocabulary – one of the long-term outcomes hypothesised in the Theory of Change. It was determined during the trial set-up stage that the underlying dimensions tested by this subtest were sufficiently proximal to those targeted by Concept Cat.

Early numeracy

Concept Cat specifically teaches early verbal concepts core to the curriculum of maths and science, with the aim of improving maths and science attainment at Key Stage 1. As such, pupils' early numeracy development will also be included as a secondary outcome. Early numeracy will be measured using the EYT ENA, which is a set of iPad-based assessment tools suitable for use with young children (ages three to six) by EY practitioners. The assessment consists of eight brief, game-like tasks that aim to capture abilities that have been found to predict later academic, social, emotional, cognitive, and life outcomes (Dawson et al., 2020). Together, these eight tasks take approximately five minutes to administer and will create the total early numeracy score.

Children's early numeracy scores are calculated by adding the number of correct responses. According to preliminary normative data reported by the EYT, scores may range from 0 to 85.[3] The ENA covers the following skill domains (Howard et al., 2022), covered by a maximum total of 85 items:

- **number sense**, which pertains to early numerical concepts and language (12 items) and rapid quantitative comparison (6 items);
- **cardinality and counting**, which refers to counting a subset of items (6 items), identifying digits and quantities (6 items), matching digits and quantities (6 items), completing number sequences (6 items), discerning the relative position of digits based on their quantity (6 items), and identifying the ordinal position of an object with respect to other objects in a line (6 items);
- **numerical operations**, which measures a child's ability to derive information from a basic, verbal mathematical problem (6 items) and solving basic numerical equations (6 items);

[3] The Early Years Toolbox team have published preliminary normative data, reported in the following link: http://www.eytoolbox.com.au/toolbox-norms.

- **spatial and measurement constructs**, which assesses a child's ability to understand spatial and measurement concepts, such as length, size, and geospatial relations (13 items); and
- **patterning**, which refers to children's ability to discern and complete increasingly complex patterns (6 items).

The composite ENA score is the sum of children's correct responses in each of these items, including those that have been marked correct due to skip rules, as described in the next paragraph. It must be noted that, while ENA items can be categorised along discrete skill domains, this evaluation will use only the composite score to estimate impact on early numeracy secondary outcome.

Skip rules make it so that children start at different points in the game based on their age, with credit given for earlier items. If the child fails on early items, they are returned to skipped items. Stop rules end the game after five consecutive incorrect responses, with subsequent unattempted trials considered incorrect (EEF, 2023b). We believe this test is suitable for the following reasons (Howard et al., 2022):

a) it measures early mathematics;

b) the mathematics measure has parallel forms with good test–retest reliability (r=0.89) and has been shown to have highly comparable results whether used by a researcher or an educator, thus making it useful for baseline assessment (see succeeding discussion below);

c) it is easy to administer using iPads; and

d) it has shown developmental sensitivity, hence minimising risk of ceiling and floor effects.

As with the primary outcome, secondary outcome testing will be administered one-to-one by independent test administrators from Elklan, who will be trained in the use and administration of the assessment and will be blinded to allocation status. These assessments will be collected only at endline owing to restrictions on the implementation timeline (i.e., need to randomise in September and allow 30-weeks for delivery).

Sample size

A total of 90 EY settings will be included in the trial, sampled from three SPHs. Sample size calculations are shown in Table 2 below.

Table 2: Sample size calculations

		Overall	EYPP
Minimum Detectable Effect Size (MDES)		0.24040	0.279279
Pre-test/ post-test correlations	level 1 (pupil)	0.75	0.75
	level 2 (setting)	0.15	0.15
Intracluster correlations (ICCs)	level 2 (setting)	0.15	0.15
Alpha		0.05	0.05
Power		0.8	0.8
One-sided or two-sided?		2	2
Average cluster size		15[4]	3
Number of schools	Intervention	45	45
	Control	45	45
	Total	90	90
Number of pupils	Intervention	675	68
	Control	675	68
	Total	1350	136

Power calculations were used to estimate the minimum sample size required to detect a set MDES (minimum detectable effect size) for this impact evaluation.

As presented in Table 2, the MDES for this investigation, given all the assumptions, is 0.247.Our calculations also assumed a 50:50 randomisation allocation to treatment and control, with the alpha level set to 0.05, power at 0.80, and an assumed average of 15 pupils per setting. In accordance with the EEF guidance, we assumed an intra-cluster correlation rate of 0.15 between settings (EEF, 2019). Whilst the CELF Preschool-2 has a published test-retest correlation of 0.95, we were more conservative with our estimates, assuming a correlation of 0.75, as this was the test-retest correlation we found in our evaluation of the Nuffield Early Language Intervention (NELI).[5]

As is standard in EEF trials, we will run a subgroup analysis on children from disadvantaged backgrounds. In EY interventions, disadvantage can either be operationalised by the number of three- and four-year-olds in receipt of Early Years Pupil Premium (EYPP) or the number of 2-year-olds eligible for the Free Early Educations Entitlement (FEEE). Given take-up of EYPP is lower for three- and four-year-olds than take up of FEEE amongst two-year-olds,[6] using EYPP as the basis for power calculations provides a more conservative estimate of MDES. We estimate that the average number of three- and four-year-olds registered for

[4] Based on the number of children that can be visited over the course of two days of testing.

[5] Nuffield Early Language Intervention (re-grant) | EEF (educationendowmentfoundation.org.uk) Nuffield Early Language Intervention (re-grant) | EEF (educationendowmentfoundation.org.uk)

[6] The Department for Education reports that 135,400 2-year-olds were registered for FEEE in 2022, whereas only 116,500 3 and 4-year-olds were in receipt of the EYPP in 2022. Source: https://explore-education-statistics.service.gov.uk/find-statistics/education-provision-children-under-5

EYPP in each setting across England is 2.4.[7] Assuming the intervention settings are representative of settings across England, we thus estimate that 360 pupils in the sample will be in receipt of EYPP within the intervention.

The above calculations do not take attrition into account. If we assume attrition at the pupil level to be at 23%[8] and setting-level attrition to be at 15%[9] [10] we have a range of potential MDES from 0.240 to 0.262, as can be seen in the table below.

	N settings	N children	MDES
At randomisation	90	1350	0.240
Setting attrition at 15%	77	1155	0.260
Pupil attrition at 23%	90	1040	0.242
Setting-level attrition at 15% and pupil-level attrition at 23%	77	889	0.262

Randomisation

This efficacy trial is designed as a two-arm cluster-randomised controlled trial, with a 50:50 allocation of settings to treatment and control groups. Randomisation will occur at the setting level to avoid potential contamination between treatment and control groups, as well as to account for the whole-class nature of Concept Cat.

Randomisation will be stratified according to region so that each region has settings delivering Concept Cat. Stratification will also be on setting type (i.e., PVI, SBS) to ensure an equal balance across each type of setting in each region. Stratifying on setting type will allow us to obtain an equal number for each type of settings in treatment and control. This is an important consideration since evidence suggests that SBS have fundamental differences compared to PVI, such as higher qualified staff (Bonetti, 2020). Having an equal number of both setting types in the treatment and control group ensures that findings from the trial are applicable to all setting types. Region will also be used as a stratification variable to ensure that the delivery team has an appropriate number of settings per trainer in each region. The regions will be: West Midlands, East Midlands, Trafford, and Everton.

In general, the number of strata should be chosen to balance the benefits of stratification with the potential drawbacks of reduced statistical power and increased complexity (Freedman & Graubard, 1978; Altman & Bland, 1999). A large number of strata can lead to smaller sample sizes within each stratum, which can reduce statistical power. On the other

[7] The Department for Education reports that 116,500 3- and 4-year-olds were in receipt of EYPP in 2022 across 47,121 providers. Source: https://explore-education-statistics.service.gov.uk/find-statistics/education-provision-children-under-5

[8] This is based on the findings of a synthesis of EEF's EY trials. Source: https://d2tic4wvo1iusb.cloudfront.net/production/documents/Early-Years-Lessons-learnt-from-EEF-trials.pdf?v=1690972141

[9] This is the reported average attrition in EEF's trials. Source: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6417459/

[10] This is not specific to EY; data on setting-level attrition in EY appears unavailable.

hand, a small number of strata may not adequately control for important confounding variables, which can lead to biased estimates of treatment effects.

Settings allocated to treatment will receive Concept Cat training and will be expected to deliver the Concept Cat programme during academic year 2023/2024, whilst those allocated to control will be expected to carry on with business as usual until the following academic year (2014/2025) when they will receive training and support to deliver Concept Cat. All settings (i.e., regardless of assignment to the treatment or control group) will be provided incentives in two tranches: £200 on completion of baseline assessments and a further £200 on completion of all endline assessments. These funds are to be used at the discretion of the setting and could be used to buy an intervention programme of their choice once the trial ends.

Randomisation will happen in the first half of Autumn half term 2023. While EEF guidance suggests collecting baseline measures before randomisation (EEF 2022), because of the tight timelines settings that have booked testing will be randomised before they have all completed testing (but after they have booked their testing). Randomisation will be in September and shared with the delivery team so they can organise training, but settings will not be informed of their allocation until they complete testing. To mitigate against potential attrition, only settings that have booked testing and shared pupil data will be eligible for randomisation.

Statistical analysis

Primary analysis

The analysis of the primary outcome (i.e., the *Basic Concepts* subtest of the CELF Preschool-2) will be done on an intention-to-treat (ITT) basis, where pupil scores will be analysed-as-randomised in a complete case analysis. This is consistent with previous EEF trials, where using ITT provides a conservative estimate of an intervention's efficacy, accounting for the effects of real-world non-compliance and attrition.

The primary outcome model will be a mixed effects random intercept model, with the *Basic Concepts* subtest of CELF Preschool-2 administered at endline as the outcome. Our target parameter (independent variable) will be a binary variable indicating treatment or control assignment at randomisation (where 1 = 'treatment', 0 = 'control'). Baseline *Basic Concepts* subtest of CELF Preschool-2 scores will be included as a control variable to ensure that prior attainment is controlled for. Since setting type and region will be used as stratification variables in randomisation, these same variables will be included as control variables. In line with EEF (2022) guidance, clustering for the mixed model will be specified at the setting level to account for the nested structure of the data and to disentangle unobserved individual-level and setting-level factors.

Secondary analysis

Our secondary analysis will comprise two distinct regression models each mirroring our primary analysis, where the mixed structure and covariates used in the model will be identical, but the outcome variable will change. As mentioned above, the secondary

outcomes being explored in this impact evaluation are scores for the *Concepts and Following Directions* subtest of CELF-Preschool-2 and the composite ENA score.

Since baseline secondary outcomes will not be collected, prior attainment will be controlled for using the *Basic Concepts* subtest of CELF Preschool-2.

Estimation of effect sizes

The impact of the Concept Cat intervention on both the primary and secondary outcomes will be quantified using Hedges' *g*, a widely used and robust effect size metric that is consistent with many previous EEF trials. Confidence intervals for these effect sizes will also be reported. The Hedges' *g* coefficients will be derived from the estimated parameters from the regression models specified above, using the following equation given in the EEF evaluator guidance:

$$ES = \frac{(\bar{Y}_T - \bar{Y}_C)_{adjusted}}{\sqrt{\sigma_S^2 + \sigma_{error}^2}}$$

Where $(\bar{Y}_T - \bar{Y}_C)_{adjusted}$ is the mean difference between the intervention and control group adjusted for baseline characteristics (i.e., the coefficient for the treatment dummy in the regression) and $\sqrt{\sigma_S^2 + \sigma_{error}^2}$ is an estimate of the population standard deviation (variance).

To assess the statistical certainty of the effect size, confidence intervals for the Hedges' *g* estimation will also be reported, as will the associated *p*-value for the treatment variable's coefficient in the regression model.

To ensure that the assumptions required for the implementation of the ordinary least squares (OLS) regression method are sufficiently met, the normality of the distribution of both primary and secondary outcomes will be assessed by mapping their distribution on separate histograms. After running the regression models specified above, *qq*-plots will be generated to assess the normality in the distribution of model residuals. Should the model residual distribution be found to be non-normal, bootstrapped estimates of the parameters will be generated using the EEF Analytics package on Stata.

Sub-group analysis

Whilst the trial does not specifically focus on the variable impact of Concept Cat on children from particular socioeconomic backgrounds, the EEF's mission to 'break the link' between family income and educational success renders this an important consideration in the impact analysis. Whilst many EEF trials taking place amongst schools with older children utilise Free School Meal (FSM) status as a binary indicator of individual deprivation, this indicator is not available for EY settings. Therefore, we will use data on EYPP eligibility to construct this binary deprivation variable. In line with EEF (2022) guidance, we will use the same model as the primary analysis, with the addition of the EYPP eligibility indicator and an interaction term combining EYPP eligibility and treatment allocation.

Further sub-group analysis will be conducted on children with special educational needs and disabilities (SEND) and children for whom English is an additional language (EAL). The SEND and EAL sub-group analyses will be identical to the EYPP sub-group analysis, whereby an interaction term between SEND status or EAL status and treatment assignment

will be added to the primary analysis model. Data on subgroups will be collected directly from settings.

Analysis in the presence of non-compliance

Whilst the primary and secondary analysis outlined above will employ a complete-case analysis, allowing the estimation of the treatment effect on an ITT basis, analysis in the presence of non-compliance will also be undertaken to calculate the impact of the intervention on the primary outcome for those who *actually* received treatment.

A continuous compliance variable will be generated using the number of Concept Cat teaching sessions attended by each pupil over the whole 30-week delivery period. This assumes that all 'compliant' pupils engage with the sessions equally. This compliance variable will then be used to calculate the Complier Average Causal Effect (CACE) through a two-stage least squares (2SLS) instrumental variable approach.

The first stage will involve modelling the compliance variable as a function of the covariates included in the primary analysis, and the results of this model will be used to calculate the 'predicted' compliance of observations based on their characteristics. The second stage will then incorporate re-running the primary analysis model but including this predicted compliance in place of treatment assignment, and then instrumenting this predicted compliance with treatment assignment. The Hedges' *g* derived from this model will similarly be calculated to provide an estimate of the CACE.

Missing data analysis

Attrition from the sample is inevitable, even with the use of incentives to retain fidelity to the intervention. Attrition across both trial arms will be explored as a basic step to assess bias. Should missingness be less than 5% overall, we will conduct a complete-case analysis of the results, regardless of any systematic pattern to the missingness observed.

Should missingness in the outcome results be higher than this 5% threshold, we will model missingness at follow-up as a function of baseline covariates, including treatment. This will enable us to gauge if the data is missing at random (MAR), where there are systematic patterns to the data which can be explained by observable characteristics in the dataset. Should this analysis reveal that the missingness is systematically associated with particular variables in the dataset, we will re-estimate the parameters of the primary outcome model using multiple imputation.

Multiple imputation involves matching observations with missing data to other observations based on particular observed characteristics, and then imputing the outcome values for these incomplete cases. The parameters of the primary analysis model are then re-estimated using the additional imputed information.

Where there are high levels of missing data, findings should also be caveated by the possibility that this missingness may be not at random (MNAR), meaning that systematic patterns in this missingness exist but are not explained by observable variables in the dataset. Whilst there is no statistical method to directly detect MNAR patterns, should high levels of attrition lead us to suspect that data is indeed MNAR, further sensitivity analyses will need to be conducted.

Implementation and process evaluation (IPE) design

The IPE will gather data in relation to the implementation of the Concept Cat programme including practitioner training, implementation of the programme in relation to business as usual, the development of practitioner knowledge and skills and engagement with families. The logic model underpins the design of the IPE. Qualitative and quantitative data will be gathered through baseline and endline surveys with practitioners and parents. This will inform understanding of change in the intervention group compared to the control group. Practitioner interviews and observations in settings will inform understanding of implementation fidelity across settings. A key data source to further understand implementation fidelity will be the routinely collected data gathered by the Concept Cat Coaches.

Research questions

IPERQ1 – How closely does the Concept Cat programme, as implemented in settings, follow the intended model (implementation fidelity), as outlined in the TiDIER framework including extended implementation for focus children? What are the barriers and facilitators to implementation and how do these differ, if at all, between setting type (PVIs/SBS)?

IPERQ2 – What, if any, adaptations have been made to the programme during implementation? Why were they made? What do they look like?

IPERQ3 – What is the nature of business as usual with regard to vocabulary instruction? How does this differ between control and intervention settings? What are the similarities/differences between setting type (PVIs/SBS)? How does programme delivery differ from business as usual?

IPERQ4 – Have practitioners attended mandatory training? To what extent have training and resources supported practitioners' ability to effectively teach Concept Cat? What is the quality of delivery [i.e., how well are different components of the intervention delivered (Humphries et al., n.d., p.6)]?

IPERQ5a – To what extent have practitioners developed their knowledge about conceptual vocabulary and skills in identifying and supporting the conceptual vocabulary development of children with higher language needs (i.e., those identified as focus children)?

> IPERQ5b – To what extent are practitioners motivated to implement, and continue to implement, Concept Cat? Is this motivation different across setting type (PVIs/SBS) and if so, why?

IPERQ6 – To what extent have settings engaged families with the programme and in what ways? Are there differences between setting type (PVIs/SBS) in the ways settings have engaged with families? How is this linked, if at all, to child outcomes?

IPERQ7 – What are the barriers and facilitators for families in home implementation of the programme, particularly for focus children, disadvantaged children and those who are EAL? What, if any, are the wider impacts on the home learning environment (HLE)?

IPERQ8 – To what extent does Concept Cat result in positive or negative unintended consequences for settings, practitioners, children, families and the HLE?

Research methods

The IPE will use mixed methods approach to collect data during Wave 1 (to pre-trial the collection of data for some elements of the IPE) and during Wave 2 – the main efficacy trial (for full details, see Table), incorporating the following:

Wave 1 – Pre-trial

Training observations will be attended by two members of the evaluation team who will attend the initial online three-hour training for lead practitioners, run by the delivery team. Training observations will take place during April 2023 and will be conducted to gain an understanding of the intervention and to support development of the training observation measures with particular focus on what fidelity implementation and quality teaching of Concept Cat should look like within the settings, especially with regards to Teach, Activate, and Review. The data collected will then be used to develop an implementation fidelity framework with a built-in quality assessment element (developed during May/June 2023) which will be used in the development and analysis of the setting observation and embedded setting observation schedules and analysis.

Monitoring data during Wave 1 will consist of three separate elements.

Coach visit to settings logs will be developed by the Concept Cat developers and will be used to gather data on input (such as the words taught within the setting) implementation quality and fidelity (such as the quality of teaching and inclusion of reviewing elements), and impact. During Wave 1, the evaluation team will ensure this log is user friendly and is capturing the data accurately. The logs will be completed by Concept Cat coaches. Two members of the evaluation team will also attend four (out of the six) face-to-face setting visits run by Concept Cat coaches (two in PVI's and two in SBS) to see how these are used by coaches during the visits. The visits will take place during May and June 2023 (two will be for visit one prior to half-term and two will be for visit two after half-term). Feedback will be given to the delivery team during the latter half of August to the first part of September.

Good practice network logs will be developed by the Concept Cat developers and will be used to gather data on attendance (as outlined below), contributions made by practitioners within the supervisions and the sharing of resources and good practice. The logs will be completed by Concept Cat coaches. During Wave 1, the evaluation team will ensure this log is user-friendly and is capturing the data accurately in a way that can be analysed. Feedback will be given to the delivery team during the latter half of August to the first part of September.

Training logs/attendance data will be used to capture data in the following ways:

1. Training attendance for lead practitioners/practitioners who have received the three/one hour training (respectively). During Wave 1, the evaluation team will monitor how well these logs are capturing the data and if any changes need to be made to the system to make it more accurate. Feedback will be given to the delivery team during the latter half of August to the first part of September.
2. Group supervision attendance for lead practitioners will be used to capture the number of group sessions attended by lead practitioners. During Wave 1, the evaluation team will monitor how well these logs are capturing the data and if any changes need to be made to the system to make it more accurate. Feedback will be

given to the delivery team during the latter half of August to the first part of September.

3. Class lists/attendance/pupil turnover data will be collected to see how settings vary in the way they capture this data. Ideally, we would like settings to collect the data in the same way. If settings are found to collect this data in different ways, the evaluation team will develop a more routine way to capture this data. Settings will be asked to send in the data they already hold but in an anonymised format (i.e., names and identifying information removed) during May 2023. Feedback will be given to the delivery team during June 2023

Coach interview schedules will be developed by the evaluation team to capture qualitative data around the use of the visit to settings logs and group supervision logs (including attendance data – points 1 and 2 above). We aim to interview two (*n*=2) coaches. The data will be used to inform the delivery team of any changes that need to be made to these logs prior to wave 2 – the efficacy trial. The coach interviews will take place during July 2023. Feedback will be given to the delivery team during the latter half of August to the first part of September.

Setting staff focus group schedules will be developed by the evaluation team to capture qualitative data around implementation of the programme. We are particularly interested to understand barriers and facilitators to implementation in different setting types (PVIs/SBS), as well as investigating the most efficient way to capture class lists/pupil turnover data (point 3 above) and how practitioners feel they can engage families in the programme. We are also interested in how settings have managed to incorporate the family engagement element into the initial implementation and how families have responded: the evaluation team had originally wanted to speak to parents directly via a parent focus group however, due to low recruitment of parents it was decided that this element should be incorporated into the setting staff focus group. The data will be used to inform the delivery team of any changes that need to be made to these logs prior to wave 2 – the efficacy trial. Focus groups (*n*=4) will be held with four to five members of setting staff. The focus groups will take place during July 2023. Feedback will be given to the delivery team during the latter half of August to the first part of September.

Incentives of £150 will be offered to settings who provide attendance data information and who take part in the focus groups. Incentives will be paid to settings through Better Communications.

The timeline for Wave 1 – pre-trial can be found in Appendix A.

Wave 2 – Efficacy Trial

Baseline practitioner survey completed by all lead practitioners (*n*=100) in intervention and control settings to collect information about existing “usual practice” such as existing vocabulary teaching practice (especially in terms of settings using Word Aware or STAR approaches), existing practice in identifying children with higher language needs and other potential moderators (such as practitioner qualifications and motivation and setting type and class size). The survey will be designed by the evaluators and will be distributed, via Qualtrics, to practitioners through email during September 2023. This aspect will address IPERQ3 and IPERQ5b with regards to business and as usual and motivation to implement. The evaluation team believe that a survey is the best way to collect a large amount of data

which will be more informative in terms of understanding usual practice. The survey will also help the evaluation team to purposely select schools to visit for observations.

Training observations (*n*=2) will be completed by two members of the evaluation team for quality assurance purposes. An observation schedule, developed by the evaluation team with support from the delivery team and information gained from the pre-trial, will be developed for the three-hour online training run by the developers (delivered to lead practitioners). A separate observation schedule will be developed by the evaluation team, with support from the delivery team, to collect data on the one-hour staff training. Both schedules will be designed to ensure the key aspects of the programme as outlined in the TIDieR model are delivered and will support in monitoring quality and fidelity. Data will also be collected from a quiz (developed by the delivery team to monitor the knowledge gained through training) distributed to practitioners and staff following training and this will also monitor how the training has been received. The delivery team will share a summary of this information with the evaluation team. The training observations will take place between September 2023 and October 2023. This aspect will address IPERQ1 with regards to implementation fidelity.

Parent Baseline survey will be created by the evaluation team on Qualtrics and distributed to settings via email. The settings will be sent the links to the surveys in the comms documents sent by the delivery team. The settings will then be responsible for distributing the survey link via email to parents. The survey will be completed by the main parent/carer for each child in both intervention and control settings (*n*=1,350). Where surveys are not completed by parents, this will be followed up by the evaluation team, via the settings. The survey will be designed to establish, at baseline, current literacy practice and relevant aspects of the HLE. This aspect will address IPERQ6 with regards to monitoring home literacy practices and the HLE (when linked with endline survey data). It will also seek to understand any differences for children who are disadvantaged, children for whom EAL and children with higher language needs. The parent baseline survey will be distributed during September 2023 to October 2023. The evaluation team suggest that a survey is the best way to collect a large amount of data which will be more informative in terms of understanding 'usual practice' in the home with regards to the HLE and literacy practices. The evaluation team will consider how this may be adapted/translated for families with lower literacy skills and families with EAL.

Monitoring data during Wave 2 will consist of three separate elements:

Coach visit to settings logs will be developed by the Concept Cat developers as part of wave 1 and will be used to gather data on input (such as the words taught within the setting) implementation quality and fidelity (such as the quality of teaching and inclusion of reviewing elements), and impact. The logs will be completed by Concept Cat coaches. Data will be used to monitor implementation fidelity. This aspect will address IPERQ1. Data will be gathered in June 2024.

Good practice network logs will be developed by the Concept Cat developers and will be used to gather data on attendance (as outlined below), contributions made by practitioners within the supervisions and the sharing of resources and good practice. The logs will be completed by Concept Cat coaches. Data will be used to monitor compliance and implementation fidelity. This aspect will address IPERQ1 with regards to implementation fidelity and IPERQ4 with regards to attendance. Data will be gathered in June 2024.

Training logs/attendance data will be used to capture data in the following ways:

1. Training attendance for lead practitioners/practitioners who have received the three/one hour training (respectively). Data will be used to monitor compliance as outlined. This aspect will address IPERQ1 and IPERQ4. Data will be gathered in October 2023.
2. Group supervision attendance for lead practitioners will be used to capture the number of group sessions attended by lead practitioners. Data will be used to monitor compliance as outlined. This aspect will address IPERQ1 and IPERQ 4. Data will be gathered in June 2024.
3. Class lists/attendance/pupil turnover data will be collected to ensure children are eligible for the intervention (i.e., attend 15+ hours per week) at baseline and to monitor how much dosage of the intervention is received by the intervention children at endline. This data will also be used to inform impact. Baseline data will be gathered following recruitment in September 2023 and endline data will be gathered in June 2024.

Setting visit observations will be conducted with intervention settings (*n*=4) and control settings (*n*=4) who will be purposely sampled depending on setting type (to have an even mix of PVI's and SBS) and the answers given in the baseline practitioner survey. The purpose of the half-day setting visits is to establish what Concept Cat looks like in settings against 'usual practice' (especially in terms of settings using Word Aware or STAR approaches). Observation schedules will be developed by the evaluation team with input from the developers. The schedules themselves will be designed to understand whether similar approaches [to Concept Cat] are being used in control settings. This aspect will cover IPERQ2 to assess any adaptations intervention settings have made to the implementation of the programme, IPERQ3 which assesses the nature of business as usual in terms of normal conceptual vocabulary instruction and normal practice in identifying children with higher language needs (which will compliment survey data), and IPERQ4 to address the quality of delivery in intervention settings. Setting visits will take place over the period of February 2024 to May 2024.

Embedded setting visit observations will be conducted in intervention settings (*n*=4) who have not taken part in the general setting observations. The purpose of these observations which will take place over a number of days is to monitor implementation fidelity to understand, on a more in-depth basis, how the different elements of the intervention (whole class, implicit play, whole class review and family) are implemented over a number of days. Observation schedules will be developed by the evaluation team with support from the developers and will be informed by training and the pre-trial observations. Settings will be purposely sampled based on setting type (an equal mix of PVIs and SBS) and answers given within the survey. This aspect will cover IPERQ1, IPERQ2 and IPERQ4 to cover implementation fidelity, adaptations to the programme, and to understand the quality of delivery and the extent to which training and resources support practitioner ability to teach Concept Cat. The embedded setting observations will take place during the period of April 2024 to June 2024, so settings have time to fully embed the programme and make any changes suited to their setting.

Practitioner Interviews will be conducted with practitioners from intervention settings (*n*=8) and control settings (*n*=4) following the setting observations and embedded setting observations. Interview schedules will be developed by the evaluation team to understand business as usual (control settings), implementation fidelity, particularly focusing on teachers' skills, resources and knowledge and how this is embedded in teaching practice, and how practitioners support those with higher language needs. The interview will also

cover barriers and facilitators to delivery as intended and adaptations and the reasons why adaptations, if any, have been made. This aspect will cover IPERQ1, IPERQ2, IPERQ4 and IPERQ5a and IPERQ5b. The interviews will take place over the period of February 2024 to June 2024.

Practitioner endline surveys will be developed by the evaluation team and distributed to all lead practitioners in control and intervention groups (*n*=90) and also to setting managers (n=100). The purpose of the survey will be to collect data to answer a number of different research questions (IPERQ1, IPERQ2, IPERQ4, IPERQ5a, IPERQ5b, IPERQ6 and IPERQ8). The survey will seek to uncover any changes in practitioner knowledge, understanding and motivation, perceived changes in parental engagement, programme adaptations, perceived impact on child attainment (especially for disadvantaged children, children with higher language needs and EAL children) and potential wider impacts and unintended consequences. The survey with setting managers will also include questions with regards to the costs of the programme and unintended consequences. The survey will be distributed in May 2024 to allow practitioners enough time to complete it prior to the end of the intervention period. The evaluation team suggest that a survey is the best way to collect a large amount of data which can be matched to baseline survey data where needed.

Parent endline surveys will be developed by the evaluation team and distributed via a Qualtrics link to settings via delivery team comms. Settings will then be responsible for distributing the link to parents. Where parents have not completed the survey, this will be followed up by the evaluation team via the settings. The endline survey will seek to establish any changes in home literacy practices and the HLE and establish, for families in intervention settings, how settings have provided information on words, additional support for children with higher language needs, barriers and facilitators to implementing the programme in the home environment, perceived gains in children's conceptual knowledge and wider language and any unintended consequences. This aspect will cover IPERQ6 to establish engagement in the programme and whether this may be linked to child outcomes; IPERQ7 to understand barriers and facilitators of home implementation and any wider impacts on the HLE; and IPERQ8 to understand any unintended consequences of the programme for the families. Parent surveys will be distributed in May 2024 to allow parents enough time to complete prior to the end of the programme. The evaluation team suggest that a survey is the best way to collect a large amount of data which can be matched to baseline survey data where needed.

The timeline for Wave 2 – Main trial can be found in Appendix B.

Analysis

The IPE has been designed to test the workings of the logic model to check whether the intervention is operating as hypothesised. Table below shows how the methods described above will answer the research questions and how they link to the implementation dimensions. In addition, Appendix G shows how the findings will be used to support or counter the logic model and its constituent elements and how it links to the implementation dimensions.

Table 3: Research questions mapped to implementation dimensions and research methods

Focus	Research Questions	Data Collection										Implementation Dimensions							
	Covered in Wave 1 and trial	Monitoring data			Observations of Training Delivery	Setting Observations	Embedded Setting Observations	Practitioner Interviews	Baseline Practitioner Survey and quiz	Endline Practitioner survey and quiz and manager Survey	Baseline Parent Survey	Endline Parent Survey	Fidelity/Adherence	Dosage	Quality	Reach	Responsiveness	Programme Differentiation	Monitoring of Control Condition	Adaptation
		Coach visits to settings	Good Practice Network	Training Logs / attendance data																
Implementation	**IPERQ1** - How closely does the Concept Cat programme, as implemented in settings, follow the intended model (implementation fidelity), as outlined in the TIDieR framework including extended implementation for focus children? What are the barriers and facilitators to implementation and how do these differ, if at all, between setting type (PVIs/SBS)?																		
	IPERQ2 – What, if any, adaptations have been made to the programme during implementation? Why were they made? What do they look like?																		
	IPERQ3 – What is the nature of business as usual with regard to vocabulary instruction? How does this differ between control and intervention settings? What are the similarities/differences between setting type (PVI's/SBS)? How does programme delivery differ from business as usual?																		
Impact on Teaching Knowledge and Practice	**IPERQ4** – Have practitioners attended mandatory training? To what extent have training and resources supported practitioners' ability to effectively teach Concept Cat? What is the quality of delivery (i.e., how well are different components of the intervention delivered? (Humphries et al (n.d., p.6))?																		

Focus	Research Questions	Data Collection											Implementation Dimensions							
	Covered in Wave 1 and trial	Monitoring data			Observations of Training Delivery	Setting Observations	Embedded Setting Observations	Practitioner Interviews	Baseline Practitioner Survey and quiz	Endline Practitioner survey and quiz and manager Survey	Baseline Parent Survey	Endline Parent Survey	Fidelity/Adherence	Dosage	Quality	Reach	Responsiveness	Programme Differentiation	Monitoring of Control Condition	Adaptation
		Coach visits to settings	Good Practice Network logs	Training Logs / attendance																
Impact on Teaching Knowledge and Practice	**IPERQ5a** – To what extent have practitioners developed their knowledge about conceptual vocabulary and skills in identifying and supporting the conceptual vocabulary development of children with higher language needs (focus children)?																			
	IPERQ5b – To what extent are practitioners motivated to implement, and continue to implement, Concept Cat? Is this motivation different across setting type (PVIs/SBS) and if so, why?																			
Impact in the home	**IPERQ6** – To what extent have settings engaged families with the programme and in what ways? Are there differences between setting type (PVIs/SBS) in the ways settings have engaged with families? How is this linked, if at all, to child outcomes?																			
	IPERQ7 – What are the barriers and facilitators for families in home implementation of the programme, particularly for focus children, disadvantaged children and those who are EAL? What, if any, are the wider impacts on the home learning environment?																			
Unintended Consequences	**IPERQ8** – To what extent does Concept Cat result in positive or negative unintended consequences for settings, practitioners, children, families and the home learning environment?																			

Table 4 summarises the range of methods that will be used to collect data for the IPE, and how it relates to the IPE research questions. The final column indicates what data will be collected. The narrative for analysis follows the table and is presented by research question to show how the data will be triangulated with impact data, where relevant, and how it will test the ToC and causal mechanisms/assumptions using a synthesised approach.

Table 4: IPE Methods overview

IPE dimension	RQ addressed	Research methods	Data collection methods	Sample size and sampling criteria	Data analysis methods
Fidelity/ Adherence	IPERQ1, 2, 4, 6	Practitioner surveys (baseline/endline)	Online questionnaires	90 practitioners (45 control, 45 intervention) 90 setting managers (45 control, 45 intervention)	Descriptive statistics; thematic analysis
		Interviews	Semi-structured interviews	8 practitioners (intervention) 4 practitioners (control)	Deductive coding; thematic analysis
		Embedded observations	Structured observations	4 intervention settings	Deductive coding; thematic analysis
		Monitoring data	Coach visit to setting logs, good practice network logs, training logs/attendance data	From all intervention settings (45)	Descriptive statistics; thematic analysis
Dosage	IPERQ4, 7	Monitoring data	Attendance at training data	From all intervention settings (45)	Descriptive statistics
		Interviews	Semi-structured interviews	8 practitioners (intervention)	Deductive coding; thematic analysis
Quality	IPERQ1, 2, 4	Embedded observations	Structured observations	4 intervention settings	Deductive coding; thematic analysis
		Training delivery observations	observation	2 sessions (by 2 members of the ET)	Descriptive statistics; thematic analysis
		Setting observations	Structured observations	4 intervention, 4 control	Deductive coding; thematic analysis
Reach	IPERQ5a, 5b	Monitoring data	Attendance at training data	From all intervention settings (45)	Descriptive statistics
		Interviews	Semi-structured interviews	8 practitioners (intervention)	Deductive coding; thematic analysis
		Practitioner surveys (baseline/endline)	Online questionnaires	90 practitioners (45 control, 45intervention) 90 setting managers (45 control, 45 intervention)	Descriptive statistics; thematic analysis
Responsiveness	IPERQ5a, 5b, 6, 8	Interviews	Semi-structured interviews	8 practitioners (intervention)	Deductive coding; thematic analysis

		Parent surveys (baseline/endline)	Online questionnaires	Approx 675 (intervention)	Descriptive statistics; thematic analysis
		Embedded observations	Structured observations	4 intervention settings	Deductive coding; thematic analysis
Programme differentiation	IPERQ1, 6, 7, 8	Embedded observations	Structured observations	4 intervention settings	Deductive coding; thematic analysis
		Interviews	Semi-structured interviews	8 practitioners (intervention)	Interviews
Monitoring of control condition	IPERQ3, 7	Setting observations	Structured observations	4 intervention, 4 control	Deductive coding; thematic analysis
		Parent surveys (baseline/endline)	Online questionnaires	1350[11](intervention and control)	Descriptive statistics; thematic analysis
		Interviews	Semi-structured interviews	8 practitioners (intervention) 4 practitioners (control)	Deductive coding; thematic analysis
Adaptation	IPERQ2,3	Embedded observations	Structured observations	4 intervention settings	Deductive coding; thematic analysis
		Practitioner surveys (baseline/endline)	Online questionnaires	90 practitioners (45 control, 45 intervention) 90 setting managers (45 control, 45 intervention)	Descriptive statistics; thematic analysis
		Interviews	Semi-structured interviews	8 practitioners (intervention)	Interviews

IPERQ1 – How closely does the Concept Cat programme, as implemented in settings, follow the intended model (implementation fidelity), as outlined in the TIDieR framework, including extended implementation for focus children? What are the barriers and facilitators to implementation and how do these differ, if at all, between setting type (PVIs/SBS)?

Measures and relation to ToC and causal mechanisms

To monitor implementation fidelity (i.e., how closely the implemented programme follows the intended model) we will bring together a number of measures. To ensure the researchers fully understand what the intended model should look like, the evaluation team will develop an observation schedule and observe training sessions during the Wave 1 pre-delivery phase to note the key elements of the programme (teach, activate, review) and, importantly, monitor how practitioners are engaging in training.

Embedded setting visits will take place in Wave 2 to monitor how the intervention is being implemented in settings focusing on the key elements of the programme with particular focus on testing underlying theory: combining implicit and explicit teaching; teaching one concept at a time; several opportunities to learn the words in depth in the everyday environment; reviewing words over time to increase retention and the concept cat methodology will lead to increased child engagement in word learning.

[11] Maximum number of parent surveys to be sent out with an estimated 30% response rate.

To collect data around barriers and facilitators to implementation we will develop a semi-structured interview schedule (also developed for the purposes of answering IPERQ2, IPERQ4 and IPERQ5) which will be designed to uncover barriers and facilitators around the key inputs and outputs outlined in the ToC. Endline surveys will also be developed to look more generally at key facilitators and barriers across the intervention group.

Synthesis of data and analysis

Training observation data from the pre-trial phase will be analysed thematically to draw out the key elements of the programme as well as draw out what implementation fidelity and 'quality' teaching will look like within the settings. This data will then be used to develop the observation schedule, with a built-in quality scale, (along with the theoretical assumptions outlined above) for the embedded setting visits (which will be used for the purposes of answering IPERQ1, IPERQ2 and IPERQ4).

Observation data will be transcribed and coded in NVivo using a mix of inductive and deductive analysis to build themes and identify patterns within the data, which relate directly to the relevant inputs and outputs outlined in the ToC. Interview and survey data will be transcribed and coded in NVivo using a mix of inductive and deductive analysis. Key barrier and facilitator themes will be analysed by setting type (PVI's/SBS) to understand any differences between settings and will be used to inform data collected on implementation fidelity. Data will be triangulated with child outcome data (collected in the impact evaluation) to support understanding of how implementation fidelity may impact on children's conceptual knowledge, and whether this differs by setting type.

IPERQ2 – What, if any, adaptations have been made to the programme during implementation? Why were they made? What do they look like?

Measures and relation to ToC and causal mechanisms

Adaptations to the programme and what and why adaptations were made will be explored through the embedded setting visits, general setting visits, practitioner interviews and practitioner endline surveys. Specifically, we will look at adaptations in relation to the key elements of the programme: implicit and explicit teaching approaches, whole class review and engagement of families.

Observation schedules for setting visits will be developed through analysis of the training observations during the pre-trial phase, Wave 1 (as outlined in IPERQ1 above), which will allow the researchers to identify adaptations and what they look like. Interviews will follow up on these changes and probe further into why the changes were made. Practitioner endline surveys will also ask more generally about adaptations, what they were, and why they were made.

Synthesis of data and analysis

All data will be transcribed (where needed) and coded in NVivo using a mix of inductive and deductive analysis. Key themes for changes in implementation, why changes were made and what the changes were will be derived from all measures and will be presented in a table in the final evaluation report. Data will be triangulated with IPERQ1 to give a broader understanding of implementation fidelity and the potential impact on child outcomes data.

IPERQ3 – What is the nature of business as usual with regard to vocabulary instruction and identifying children with higher language needs? How does this differ between control and intervention settings? What are the similarities/differences

between setting type (PVIs/SBS)? How does programme delivery differ from business as usual?

Measures and relation to ToC and causal mechanisms

To understand 'business as usual' with regard to vocabulary practice (especially in terms of settings using Word Aware or STAR approaches), identify children with higher language needs (including data on children who are receiving speech and language support), and to capture heterogeneity in usual practice, we bring together data from practitioner baseline surveys and control group observation data.

Baseline surveys will be designed to establish similarities and differences between business as usual and Concept Cat. Specifically, we would seek to understand usual conceptual vocabulary instruction as well as ascertain if any resources used as part of the intervention (i.e., Word Aware book) are already being used within settings (control and intervention) and the extent to which they may be used.

We also seek to understand how children with higher language needs are identified in usual practice and we seek to understand differences and similarities with how focus children are identified within the Concept Cat Programme. Appendix C outlines the identification of children with higher needs within the Concept Cat programme. Control group settings will be purposely sampled based on baseline survey answers (2 PVIs, 2 SBS) to understand similar business as usual practice and different business as usual practice. Intervention group observations will be conducted to review the ToC's inputs (for pupils and families) and outputs (for teachers, pupils and families) and to further understand differences between intervention and control settings.

Synthesis of data and analysis

Baseline teacher survey data will be mainly analysed descriptively, and we will separate data from PVIs and SBS to look at similarities and differences between setting type. To ensure data can be synthesised, qualitative data obtained from the survey and observation data will be coded in NVivo using a mix of inductive and deductive analysis to build themes and identify patterns within the data which relate directly to the ToC's inputs. Data will be presented comparatively to look at differences between control and intervention settings and differences between PVIs and SBS.

IPERQ4 – Have practitioners attended mandatory training? To what extent have training and resources supported practitioners' ability to effectively teach Concept Cat? What is the quality of delivery [i.e., how well are different components of the intervention delivered (Humphries et al (n.d., p.6)]?

Measures and relation to ToC and causal mechanisms

Practitioners' attendance at training will be gathered as part of routinely collected data from the developers and will provide information relating to the amount of training received in settings. Data from the practitioner quiz – which is completed after receiving training – will be used to understand how useful practitioners have found the training.

Data on support received from coaches and peers will also be collected via coach visit to setting logs and will inform the level of compliance in terms of support received from Concept Cat Coaches (with six regular visits deemed as high compliance, five as medium compliance and four as low compliance) and good practice network logs (with attendance at

five group supervisions deemed as high compliance, four as medium compliance and three or less as low compliance).

The extent to which training and resources have supported the practitioner's ability to effectively teach Concept Cat will be explored through coach visits logs and group supervision logs, which will be used to assess changes that have been made in the settings and in practitioner practice over the course of the intervention, as well as what the changes look like.

Good practice network logs will detail sharing of good practice and resources and positive contributions made in the sessions. Practitioner interview questions will be designed to explore further how training and materials have supported delivery and survey questions will also probe into the practitioner's own perceptions of how well training and resources supported them in delivery. The quality of delivery will be assessed through both embedded observations and general observations within intervention settings. Observation schedules will be developed through training observation data (as outlined above in IPERQ1) to be indicative of quality of practice.

Synthesis of data and analysis

Attendance data will be descriptively analysed to assess the level of training undertaken across settings and the usefulness of the training (i.e., quiz data). Coach visit logs and good practice network logs will also be analysed descriptively to monitor the level of compliance and monitor the level of support received. Data from the coach visit logs and interview data will be separately transcribed and coded in NVivo using a mix of inductive and deductive analysis to build themes and identify patterns within the data which relate directly to the ToC's short-term outcomes.

Survey data will be descriptively analysed and will be cross-referenced with training logs and interview data. Observation data will be coded in NVivo using a mix of inductive and deductive analysis to build themes and identify patterns within the data which relate directly to the quality of practice outlined. Data will be triangulated with child outcome data (collected in the impact evaluation) to support understanding of how training and quality of teaching practice may impact on children's conceptual knowledge.

IPERQ5a – To what extent have practitioners developed their knowledge about conceptual vocabulary and skills in identifying and supporting the conceptual vocabulary development of children with higher language needs (i.e., those identified as focus children)?

Measures and relation to ToC and causal mechanisms

Practitioners' development of knowledge to identify and support those with higher language needs will be assessed through practitioner endline surveys and practitioner interviews. Practitioner surveys will be used to establish the level of knowledge and understanding gained throughout the programme with particular reference to: (a) understanding the importance of conceptual vocabulary development; (b) having the skills and resources to effectively incorporate repetition of words into everyday practice through implicit and explicit teaching methods; and (c) the development of skills in identifying and supporting conceptual language development in children with higher language needs. Interview questions will probe more deeply into what the change looks like within the settings and will test the causal assumption that training, coaching, and modelling to staff will lead to change in understanding and practice.

Synthesis of data and analysis

Practitioner survey data will be analysed descriptively and will be presented to show change over time in knowledge and skills and differences/similarities between intervention and control settings. Interview data will be transcribed and coded in NVivo using a mix of inductive and deductive analysis to build themes and identify patterns within the data which relate directly to the ToC's short-term outcomes. Interview data will be cross-referenced with practitioner survey data.

IPERQ5b – To what extent are practitioners motivated to implement, and continue to implement, Concept Cat? Is this motivation different across setting type (PVIs/SBS) and if so, why?

Measures and relation to ToC and causal mechanisms

To understand motivations to implement Concept Cat, practitioners will be asked to rate their level of motivation and the reasons for their level of motivation within the practitioner baseline and endline survey. This will allow the evaluation team to assess the causal assumption that teachers are motivated to implement and sustain the changes to teaching practice.

Synthesis of data and analysis

Rated data collected from the baseline and endline surveys will be analysed descriptively to see if any changes in the level of motivation to implement and continue to implement Concept Cat changes over the course of the intervention. We will also split this data to assess any differences between setting type (PVIs/SBS).

Qualitative data on the reasons for the rated level of motivation will be coded in NVivo using a mix of inductive and deductive analysis to build themes and identify patterns within the data. We will also split this data to assess any differences between setting type (PVIs/SBS).

IPERQ6 – To what extent have families engaged with the programme and in what ways? Are there differences between setting type (PVIs/SBS) in the ways families have engaged? How is this linked, if at all, to child outcomes, particularly for those who are EAL, disadvantaged or have higher learning needs?

Measures and relation to ToC and causal mechanisms

Data relating to family engagement in the programme will be collected from both families and practitioners through the endline surveys. Practitioners will be asked to provide information on how concept words and activities were shared with families and their opinions of whether families were engaged with the programme at home. Families will be asked to provide information on how well the settings shared concept words and activities, to test the inputs outlined in the ToC, and how often they did the activities with their children, to test the short-term outcomes outlined in the ToC. The parental survey will also ask questions about the home literacy and learning environment both at baseline and endline to enable the evaluators to establish changes over time in the HLE and test the causal assumption that talk activities at home support vocabulary acquisition.

Synthesis of data and analysis

Survey data will be analysed descriptively to establish the perceived level of engagement with the programme in the home. Data on how well the concept words and activities were shared will be rated and used to establish whether the way in which data was shared

impacted on the level of engagement in PVIs and SBS. This information will help to inform the developers as to the best ways for settings to engage parents in the programme and will also be used to inform data presented on changes in the HLE. Data on home literacy practices and the HLE will be analysed descriptively to assess changes over time in the intervention settings (and whether this was moderated by setting type or the way the concept words and activities were shared), as well as differences and similarities between control and intervention settings at baseline and endline. The data will be triangulated with child outcome data, collected as part of the impact evaluation, to assess whether talk activities in the home supported concept vocabulary acquisition, particularly for those who have EAL, suffer the effects of being disadvantaged or who have higher learning needs.

IPERQ7 – What are the barriers and facilitators for families in home implementation of the programme, particularly for focus children, disadvantaged children and those who have EAL? What, if any, are the wider impacts on the HLE?

Measures and relation to ToC and causal mechanisms

Data on barriers and facilitators to home implementation of the programme will be gathered through parent/carer endline surveys. Questions will be designed around the family elements outlined in the ToC to develop an understanding of which elements (inputs/outputs/short-term outcomes) families found helpful/unhelpful. Families will also be asked about wider impacts on the HLE, such as filtration of the programme to siblings and engagement from other family members, amongst others.

Synthesis of data and analysis

Data on barriers and facilitators will be analysed descriptively and will be helpful to inform the developers for future programme implementation and evaluation. Data will be triangulated with data collected as part of the impact evaluation on EYPP to understand and report on barriers and facilitators specific to families with low SES. Data will also be triangulated with impact evaluation data pertaining to EAL status to understand and report on barriers and facilitators specific to families who are classed as EAL.

IPERQ8 – To what extent does Concept Cat result in positive or negative unintended consequences for settings, practitioners, children, families, and the HLE?

Measures and relation to ToC and causal mechanisms

Data on positive or negative unintended consequences will be gathered through family and practitioner baseline surveys. Questions will be designed around potential changes in the setting and the home, and participants will be asked to rate (on a scale of one to five) which best applies to them. For practitioners, questions will centre on changes in staff turn-over, setting environment, and cost and time of implementing the programme. For families, questions will focus on changes in child behaviour or sibling behaviour, changes in the home environment and changes in time spent doing activities with the other children or family members in the household.

Synthesis of data and analysis

Data will be analysed descriptively to be informative for future programme implementation and evaluation.

Cost evaluation design

Data on implementation costs will be collected through endline surveys with practitioners in both control and treatment settings. In addition, we will account for hidden savings and costs that may surface through surveys and interviews with practitioners. These hidden savings and costs, if they exist, will be monetised using market estimates and added to explicitly reported costs to determine a per-pupil-school-year cost, as recommended by EEF's Cost Evaluation Guidance

Cost data from control settings will be used to compare the cost of implementing Concept Cat against business as usual, which could correspond to the costs associated with implementing other programmes that are similar in scope to Concept Cat. Performing this comparison also aligns with EEF's 2023 Cost Evaluation Guidance.[i] We note that business as usual is likely to differ across control settings; as such, we will perform sensitivity analyses to account for heterogeneity.

The cost evaluation will look at both direct and indirect costs incurred by implementing Concept Cat and those incurred by implementing similar programmes. These direct and indirect costs include but are not limited to: (a) time away from teaching due to participation in training and other programme activities; (b) staff cover for teaching staff participating in out-of-setting programme-related activities; (c) prices of instructional materials; and (d) additional staff workload required to run the programme.

Ethics and registration

This evaluation will be registered on the International Standard Randomised Controlled Trial Number (ISRCTN) Registry. Once the evaluation has been registered, we will update this protocol with the corresponding registration number.

The procedures described in this protocol are in line with the ethical standards of RAND Europe and the University of York. They have been reviewed and approved by both RAND US' Human Subjects Protection Committee (HSPC) on 14 March 2023 and the University of York's Education Ethics Committee (EEC) on 7 March 2023.

Consent to participate in the intervention and evaluation will be obtained from parents or legal guardians, who act as decision-makers for individual pupils. In the interest of informed consent, settings will provide parents and legal guardians with information sheets and withdrawal forms. If parents or legal guardians decide to withdraw their child from the intervention, evaluation, or both, they may return the withdrawal form. Parents or legal guardians may withdraw their children at any time throughout the intervention and evaluation. If a child's participation is withdrawn, the delivery and evaluation teams will not collect data from the child or will delete any data previously collected, as needed.

The University of York will provide and collect consent forms from setting staff who participate in observations, interviews, or focus group discussions for the IPE. In addition, surveys with parents and practitioners will include a privacy notice indicating to respondents that their participation is voluntary and that they can choose to withdraw at any time without penalty.

No member of the evaluation team has any conflict of interest with respect to the intervention or evaluation.

Data protection

The Data Protection Impact Assessment for this project was approved and signed off by RAND Europe's Data Protection Officer on 19 January 2023.

Several teams are involved in controlling and processing data. RAND Europe and University of York will act as co-controllers, with Better Communication CIC and Elklan acting as processors. Further details on this are outlined in the Data flow diagram in Appendix D.

As part of the evaluation, Elklan will collect information from settings about all the children that take part in Concept Cat. RAND Europe will also ask Elklan to collect data on assessed child outcomes. Settings and Elklan will provide this information by using an Excel data collection form provided by the evaluation team. This form will be shared via secure file transfer (i.e., Egress).

The University of York will ask Better Communication CIC (the delivery team) to collect information from settings about key staff. Settings will provide this information by using an Excel data collection form. This form will be shared via secure file transfer (such as Egress). The University of York will use the data provided by settings on key staff to invite them to take part in a short survey and/or interview. RAND Europe may also use this data in its research analysis to understand the impact of the programme.

At the end of the study, RAND Europe will submit the data in pseudo-anonymised format to the Office for National Statistics Secure Research Service (ONS SRS) for archiving in the EEF data archive. This data will only be identifiable to the DfE and may be matched to the National Pupil Database (NPD) and other administrative data during subsequent research. The EEF and DfE will act as data controllers for the archive, along with contractors appointed to manage the archive.

The legal basis for RAND Europe is legitimate interests, as detailed in Article 6(1)(f) of the UK GDPR. The legal basis for processing your child's special category data[12] is because it is necessary for archiving purposes in the public interest, scientific, or historical research purposes, as detailed in Article 9(2)(j) of the UK GDPR. To ensure that all processing is fair and lawful, RAND Europe have also completed a Legitimate Interest Assessment and a Data Protection Impact Assessment, in addition to the completion of an application to the RAND internal review board for ethical approval. RAND Europe will process only what is required to meet these legal bases and will ensure security and safeguards are in place to protect the information.

The legal basis for the University of York, EEF, and DfE is where it is necessary for the performance of a task carried out in the public interest as set out in Article 6(1)(e) of the UK GDPR. The specific legislation which allows this is Section 10 of the Education Act 1996.

[12] 'Special category' data is personal data that needs more protection because it is sensitive, for example health or ethnicity data.

The legal basis for processing special category data is for reasons of substantial public interest as detailed in Article 9(2)(g) of the UK GDPR.

We take information security extremely seriously and we have appropriate technical and organisational measures to protect personal data and special category data. Access to information is restricted on a need-to-know basis and security arrangements are regularly reviewed to ensure their continued suitability. The evaluation team will collect and store all personal and special category data in accordance with the Data Protection Act (2018) and UK GDPR requirements. No personal information collected as part of this study will be transferred outside of the European Economic Area (EEA).

All individually identifiable data held by RAND Europe will be destroyed one year after the end of the study (2026). All individually identifiable data held by the University of York will be destroyed 5 years after the end of the study (2030). Data in the EEF's archive in the ONS SRS will include data only individually identifiable to the Department for Education (DfE), the government department responsible for children's services and education, and is kept indefinitely for the purposes of future research. Anonymous data will be kept indefinitely by the University of York.

Personnel

DELIVERY TEAM			
Name	**Institution**	**Role**	**Responsibilities**
Marie Gascoigne	Better Communication CIC	Director	Organise data sharing agreements with settings Conduct training and support activities for Concept Cat implementation in settings, including the delivery of one-to-one coaching sessions Recruitment of settings and ongoing liaison Recruitment of Concept Cat Coaches and organisational support
Stephen Parsons		Associate – CC Subject Expert	
Anna Branagan		Associate – CC Subject Expert	
Victoria Riley		Associate – Deputy for Marie Gascoigne	
Bibiana Wigley		Associate – Lead for setting recruitment and liaison	
Michele Seidler		Associate - Operational resources and people	

EVALUATION TEAM			
Name	**Institution**	**Role**	**Responsibilities**
Elena Rosa Brown	RAND Europe	Project Lead	Act as schools' first point of contact for questions about the evaluation Provide oversight and direction regarding evaluation design and methodology

<table>
<tr><th colspan="4">EVALUATION TEAM</th></tr>
<tr><th>Name</th><th>Institution</th><th>Role</th><th>Responsibilities</th></tr>
<tr><td>Miguel Subosa</td><td>RAND Europe</td><td>Project Manager</td><td>Ensure timely delivery of evaluation activities and outputs, including coordination with the testing partner

Manage the drafting of documents required for data protection, ethical approval, and trial registration, including the study protocol

Develop data collection instruments for the impact evaluation

Manage the development of the final evaluation report and all interim outputs</td></tr>
<tr><td>James Merewood</td><td>RAND Europe</td><td>Analyst (Statistician / Economist)</td><td>Develop the Statistical Analysis Plan

Conduct randomisation of schools to the treatment and control group

Lead and oversee the statistical analysis of CELF Preschool-2 and Early Numeracy Assessment data for the impact evaluation</td></tr>
<tr><td>Fin Oades</td><td>RAND Europe</td><td>Research Assistant</td><td>Provide research support for statistical analysis, including: cleaning up datasets; writing code for statistical analysis; and running statistical tests</td></tr>
<tr><td>Louise Tracey</td><td rowspan="2">University of York</td><td>IPE Project Lead</td><td rowspan="2">Conduct baseline and endline surveys with setting staff

Conduct setting visits to a selection of settings during the study

Develop data collection instruments for the IPE

Collect data required for the IPE

Analyse data collected for the IPE</td></tr>
<tr><td>Erin Dysart</td><td>IPE Project Manager</td></tr>
<tr><td>Alex Hall</td><td>Elklan</td><td>Testing Partner</td><td>Collect data on participating children from settings

Coordinate administration of the CELF Preschool-2 at baseline and endline

Coordinate administration of the Early Numeracy Assessment at endline

Transmit results of baseline and endline testing to RAND Europe</td></tr>
</table>

Risks

Risks	Assessment (Likelihood / Impact)	Mitigation strategy	Impact post-mitigation
Recruitment failure	Likelihood: Moderate Impact: High	This can be mitigated by regular dialogue over any recruitment issues and ensuring that the design incorporates minimal burden to settings. Timelines will be discussed and agreed well in advance to ensure there is adequate time for all activities.	Low
Attrition	Likelihood: Moderate Impact: High	Settings will be given clear information about participation before signing up. There are also incentives for participation in testing, coupled with low test-burden (i.e., completed by external testers). Finally, as wait-listed design the hope is that attrition will be minimised as all settings interested in Concept Cat will eventually have access to the programme.	Moderate
Low participation rates in data collection (testing and IPE)	Likelihood: Moderate Impact: High	Allowing sufficient data collection window, with real-time monitoring of response rates to allow for targeted reminders to be sent. Piloting measures and data collection tools to understand how to reduce burden on settings. We will also ensure that the survey can be completed on phones to make it easier for practitioners and parents who may not have easy access to a computer.	Low
Small number of EYPP children for analysis	Likelihood: Moderate Impact: Moderate	The delivery team will aim to recruit settings in areas of high deprivation to support analysis of EYPP children in both the impact and IP evaluations.	Low
Cross-contamination	Likelihood: Low Impact: High	Setting-level randomisation. Given one of the key elements of the programme are the resources we also do not see how practitioners moving between settings will automatically lead to contamination (as they would also need to bring the resources with them). Settings that are part of a nursery chain will be considered on a case-by-case basis to assess whether cross-	Low

Risks	Assessment (Likelihood / Impact)	Mitigation strategy	Impact post-mitigation
		contamination is a threat. As a check for contamination, information about all programmes used, whether Concept Cat or comparable programmes will be collected in surveys and (if necessary) factored into the analysis.	
Quality of reporting	Likelihood: Moderate Impact: Moderate	Applying RAND Europe's QA processes, including expert review. Project Leader has considerable experience using EEF reporting standards.	Low
Lack of coordination between RE and UoY teams	Likelihood: Moderate Impact: Moderate	Teams to attend initial meetings and agree on roles and responsibilities at the outset. Regular contact between key team members from each organisation.	Low
Evaluation team members absence or turn-over	Likelihood: Moderate Impact: Low	The team can be supplemented by researchers with experience in evaluation from the larger RAND Europe and University of York pool. All RE staff have a three-month notice period to allow sufficient time for handover.	Low

Timeline

Table 5: Timeline

Dates	Activity	Staff responsible/ leading
January 2023 – September 2023	Recruitment of settings	Better Communication CIC
September 2023 – October 2023	Baseline assessment (CELF Preschool-2)	Alex Hall (Elklan)
September 2023	Randomisation of settings into control and treatment groups	James Merewood (RAND Europe)
	Informing settings and Better Communication CIC of randomisation allocation	Miguel Subosa (RAND Europe)
	Administration of baseline surveys with parents and practitioners	Louise Tracey & Erin Dysart (University of York)
	Training delivery	Better Communication CIC
	Training observations	Louise Tracey & Erin Dysart (University of York)
	Delivery of Concept Cat	Better Communication CIC
	Administration of endline surveys with parents and practitioners	Louise Tracey & Erin Dysart (University of York)
	Endline assessment (CELF Preschool-2 and Early Numeracy Assessment	Alex Hall (Elklan)
January 2024 – May 2024	Setting visits	Louise Tracey & Erin Dysart (University of York)
	Practitioner interviews	Louise Tracey & Erin Dysart (University of York)
June 2024 – July 2024	Administration of Endline surveys for practitioners and parents	Louise Tracey & Erin Dysart (University of York)
July 2024 – November 2024	Analysis and drafting	and University of York

References

Altman, D. G., & Bland, J. M. (1999). Treatment allocation by minimisation. BMJ, 318(7191), 756-759.Bonetti, S. (2020). *Early years workforce qualifications and chidlren's outcomes*. Education Policy Institute. https://epi.org.uk/publications-and-research/early-years-qualifications-and-outcomes/

Bennett, J., Hanley, P., Abrahams, I., Elliott, L., & Turkenburg-van Diepen, M. (2019). Mixed methods, mixed outcomes? Combining an RCT and case studies to research the impact of a training programme for primary school science teachers. *International journal of science education*, *41*(4), 490-509. doi: 10.1080/09500693.2018.1563729

Clarke, P., Snowling, M., Truelove, E., & Hulme, C. (2010). Ameliorating Children's Reading-Comprehension Difficulties : A Randomized Controlled Trial. *Psychological Science*. https://doi.org/10.1177/0956797610375449

Dawson, A., Stokes, L., Huxley, C., Runge, J., Takala, H., Manzoni, C., Hudson-Sharp, C., & Williams, C. (2020). *Early Years Toolbox: Pilot Report*. https://educationendowmentfoundation.org.uk/public/files/Early_Years_Toolbox_Report_(final).pdf

EEF. (2019). *Lesson learnt from EEF Early Years trials: Recommendations for evaluators.* https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/EY_lessons_learnt.pdf

EEF. (2022). *Statistical analysis guidance for EEF evaluations.* https://d2tic4wvo1iusb.cloudfront.net/documents/evaluation/evaluation-design/EEF-Analysis-Guidance-Website-Version-2022.14.11.pdf?v=1681842352

EEF. (2023a). *Early Years Measures database*. https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluation/eef-outcome-measures-and-databases/early-years-measures-database-2

EEF. (2023b). *Early Years Toolkit: A summary of the best available evidence on key areas for learning and development.* Retrieved 09/02/2023 from https://educationendowmentfoundation.org.uk/education-evidence/early-years-toolkitFernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental science,* 16(2), 234-248.

Fricke, S., Bowyer-Crane, C., Haley, A., Hulme, C., & Snowling, M. (2012). Efficacy of language intervention in the early years. *Journal of Child Psychology and Psychiatry, 54*(3), 280-290. https://doi.org/10.1111/jcpp.12010

Freedman, L. S., & Graubard, B. I. (1978). Sample size for studying intermediate endpoints within intervention trials of observational studies. American Journal of Epidemiology, 108(3), 240-248.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H Brookes Publishing.

Hopkins, T., Harrison, E., Coyne-Umfreville, E., & Packer, M. (2022). A pilot study exploring the effectiveness of a whole-school intervention targeting receptive vocabulary in the

early years: Findings from a mixed method study involving students as part of a practice-based research placement. *Child Language and Teaching Therapy, 38*(2), 212-229.

Howard, SJ., Neilsen-Hewett, C., de Rosnay, M., Melhuish, EC., & Buckley-Walker, K. (2022). Validity, reliability, and viability of pre-school educators use of early years toolbox early numeracy. *Australasian Journal of Early Childhood*, 47(2), 92 - 106.

Law, J., Charlton, J., Dockrell., J., Gascoigne, M., McKean, C., & Theakston, A. (2017). *Early Language Development: Needs, provision, and intervention for preschool chidlren from socio-economically disadvantaged backgrounds.* EEF. Retrieved 27/06/2023 from https://d2tic4wvo1iusb.cloudfront.net/documents/guidance/Law_et_al_Early_Language_Development_final.pdf?v=1684346130

LeFevre, J. A., Skwarchuk, S. L., Smith-Chant, B. L., Fast, L., Kamawar, D., & Bisanz, J. (2010). Home numeracy experiences and children's math performance in the early school years. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 42(2), 177-182.

Marulis, L., & Neuman, S. (2012). The effects of vocabulary interventions on young children's word learning: A meta-analysis. *Review of Educational Research, 80*(3), 300-335. https://doi.org/10.3102/0034654310377087

Scarborough, H. (2009). Connecting Early Language and Literacy to Later Reading (Dis)abilities: Evidence, Theory, and Practice. In F. Fletcher-Campbell, J. Soler, & G. Reid (Eds.), *Approaching Difficulties in Literacy Development: Assessment, Pedagogy and Programmes*. SAGE Publications Ltd.

Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., & Newcombe, N. S. (2014). Finding the missing piece: Blocks, puzzles, and shapes fuel school readiness. *Trends in Neuroscience and Education*, 3(1), 7-13

Appendix A: Wave 1 (pre-trial IPE timeline)

Task	March 2023	April 2023	May 2023	June 2023	July 2023	August 2023	Sept 2023
Wave 1 protocol write up							
Attend Training (3-hour session)		19th April					
Attend 4 face-to face coach visits to settings			2 prior to half-term (visit 1)	2 after half-term (visit 2)			
Develop coach interview schedule							
Develop setting staff focus group schedule							
Develop parent focus group schedule							
Collect class list templates							
Coach interviews							
Staff focus groups							
Parent focus groups							
Feedback on class list templates							
Development of implementation and quality framework							
Feedback back on coach visit to settings logs							
Feedback on group supervision logs							
Feedback on coach interviews							
Feedback on setting staff focus groups							
Feedback on parent focus groups							

Appendix B: Wave 2 (main IPE timeline)

Task	Jan Feb 23	March April 23	May June 23	July Aug 23	Sept Oct 23	Nov Dec 23	Jan Feb 24	March April 24	May June 24	July Aug 24	Sept Oct 24	Nov Dec 24
Write IPE												
ISRCTN registration												
Develop pre-trail materials (see pre-trial timeline)												
Develop baseline practitioner survey												
Develop training observation schedule (3 hour and 1 hour)												
Develop baseline parent survey												
Develop implementation fidelity and quality framework												
Develop compliance spreadsheet (including attendance data)												
Develop setting observation schedule												

Develop embedded setting observation schedule												
Develop practitioner interview schedule												
Analysis of baseline data												
Develop endline practitioner survey												
Develop endline parent survey												
Attend training observations												
Distribute baseline practitioner survey												
Distribute baseline parent survey												
Setting observations												
Embedded setting observations												
Carry out practitioner interviews												
Distribute endline practitioner survey												
Distribute endline parent survey												

Collect compliance data					Training attendance data							
Analysis of endline data												
Merge analysis of baseline and endline data												
Writing of final report												

Appendix C: Identification of children with higher needs within Concept Cat programme

All children will be learning about the selected words but ‘focus children’ will get a bit more support.

Use the flowchart on the next page to help you pick the right children. Your Concept Cat Coach will help you choose the right children.

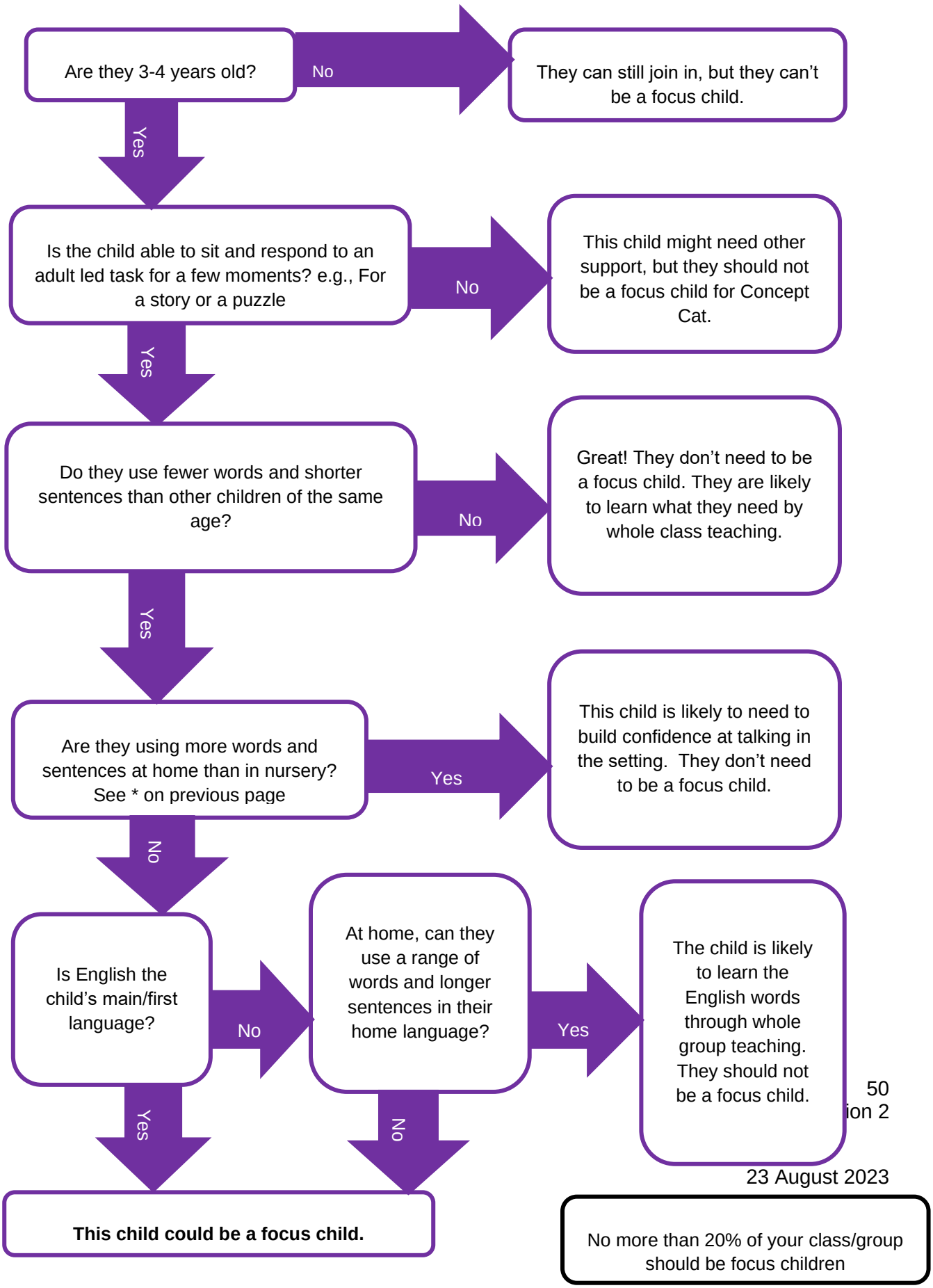
Ideally you want no more than 20% of your group to be focus children.

***** If a child isn’t speaking much in your setting it is good to talk to the family to see how the child is talking at home. The sorts of questions to ask are:

- What is your child’s longest sentence?
- What is their most advanced/grown up word?
- Do they use words such as ‘big, under, more?’

Think about how this compares with the other children of a similar age.

Focus children



50

ion 2

23 August 2023

No more than 20% of your class/group should be focus children

Focus children that attend 5 days a week

This is the additional intervention that the children have. The focus children do everything with the whole group **as well** as this.

Monday	Repeat: gesture/sign, symbol, say it and song (1) Repeat: Concept Cat story or video (2)
Tuesday	Repeat a teach activity (3) Extra contact with families of focus children: • Face to face or by phone • Tell them about the word and how it has been taught in school/nursery • Talk about the suggested activity sent on Monday. • Encourage to send in a photo for the Concept book. Key message: repeat the word, experience the word, use the word in sentences
Wednesday	Go to the Concept Cat house, tell the Concept Cat story again (2) and model with objects
Thursday	Picky puppet with current word. Use the word with children in an activate activity (4)
Friday	Picky puppet with current word. Word bag with previous week's words.

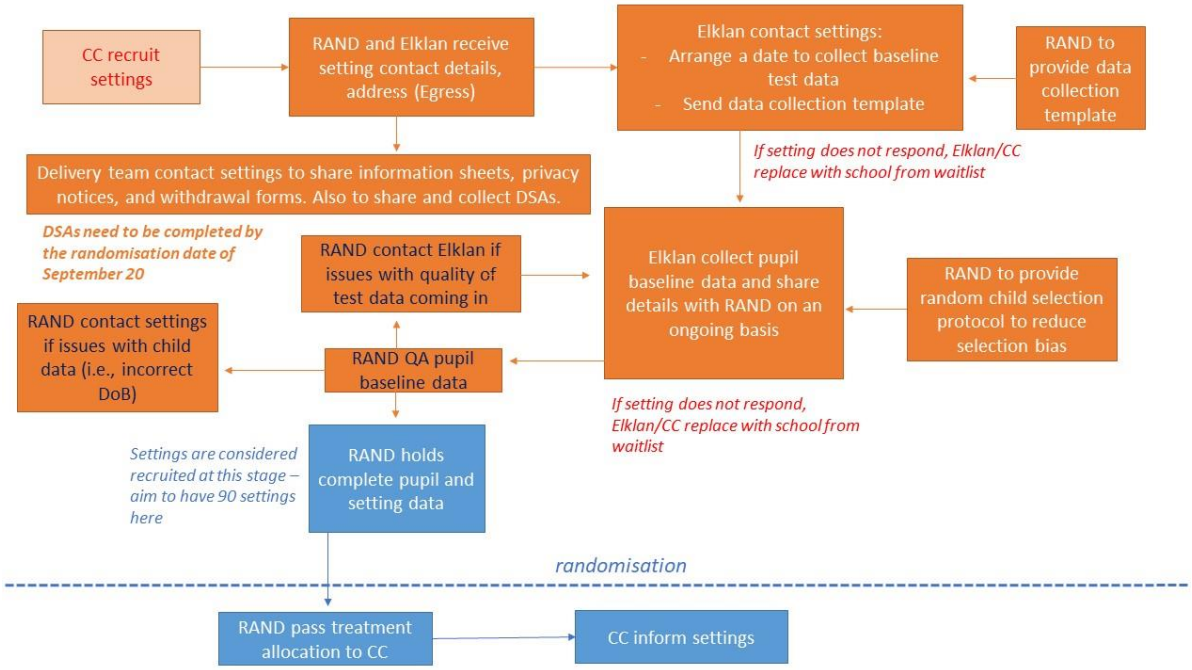
Focus children that attend 3 days a week

This is the additional intervention that the children have. The focus children do everything with the whole group **as well** as this.

Day 1	Repeat: gesture/sign, symbol, say it and song (1) Repeat: Concept Cat story or video (2)
Day 2	Repeat a teach activity (3) Go to the Concept Cat house, tell the Concept Cat story again (2) Extra contact with families of focus children: • Face to face or by phone • Tell them about the word and how it has been taught in school/nursery • Talk about the suggested activity sent on Monday. Key message: repeat the word, experience the word, use the word in sentences
Day 3	Picky puppet with current word. Take objects from the word bag. Talk about this week's concept and 2 other concepts. (5)

*Numbers relate to the steps in the teaching sequence in the WA2 book

Appendix D: Data flow diagram



Appendix E: Privacy notice

1. Background

The Education Endowment Foundation (EEF) supported by the Department for Education (DfE) is funding a study to evaluate a programme designed to improve early years language development.

Concept Cat is a whole class teaching methodology for early verbal concepts (part of the widely used book 'Word Aware 2') and delivered by authors Stephen Parsons and Anna Branagan, working with a partner organisation (Better Communication CIC).

The Concept Cat approach teaches children concepts such as 'first', 'wide' and 'empty'. This is designed to be taught in a structured and engaging manner with explicit teaching of vocabulary followed by implicit teaching in play-based learning. The teaching process includes staff acting out a scripted story with a toy cat (that we will send to each participating setting). Small changes are made to the environment so that children have chances to experience the new word an increased number of times. For instance, if the target word for the week was 'empty', then sand and water trays would be out that week. Families are engaged with simple home activities which will be available in a number of community languages.

This sequence is designed to be accessible to a wide range of children, including those with limited language. One word is taught per week, allowing opportunity for deep understanding to develop. Rather than general vocabulary, Concept Cat specifically teaches early verbal concepts (such as 'before', 'early' and 'through') core to the curriculum of maths and science, with the ultimate aim of improving maths and science attainment at Key Stage 1.

2. About the research study

RAND Europe and the Department of Education at the University of York are the independent evaluators of the Concept Cat programme.

The aim of the evaluation is to understand whether Concept Cat has had an impact on child outcomes (the 'impact' evaluation), which will be achieved by comparing data from children receiving the intervention to similar children that did not receive the intervention within a given timeframe. This will help us understand if and how the Concept Cat programme makes a positive impact on children. The intended outcomes of the programme include: early language development and early numeracy.

The gathered data is therefore expected to make an important contribution to understanding what works in early language development and pupil attainment, as well as effective teaching strategies to build conceptual vocabulary.

This privacy notice is for the parents/guardians of children in Early Years settings which are participating in the Concept Cat evaluation. It sets out the ways in which RAND Europe and the University of York, use, store, and share your data and your child's data. It also sets out how long we keep this data and what rights you have in relation to this data under the General Data Protection Regulation (GDPR).

3. Who is involved in the research study?

The evaluation is being funded by the EEF, an organisation that funds research into education, supported by the DfE. RAND Europe and the University of York are doing the evaluation research (the 'evaluation team'). RAND Europe and the University of York will be supported by Elklan, who will collect data on child outcomes through standardised assessment tools. Elklan's trained professionals will come in to work with children, in coordination with the classroom teacher. This means that they will not have direct contact with your child without the presence or awareness of the teacher or a staff member from within the setting. Thinking Talking have designed the programme and Better Communication CIC are responsible for its delivery (the 'delivery team').

4. Where do we get your data from and what data do we collect?

The evaluation team collect personal data[13] about your child for the purposes of conducting the study, to evaluate the impact and effectiveness of the Concept Cat programme, as well as to prepare a report about the study.

Your child's Early Years setting will share the following information about your child with University of York:

- Full name
- Date of birth
- Gender
- Home postcode
- Early years setting postcode (or Local Authority Establishment number and the school's Unique Reference Number, if school based)

[13] By 'personal data', we mean any information about you or your child which could be used to identify you, such as your name.

- Attendance data (number of hours and frequency)

This information will be used for future data linkage to the National Pupil Database (NPD)

- Early Years Pupil Premium (EYPP) status
- English as an Additional Language (EAL) status
- Special Educational Needs (SEN) status

This data will be matched to the following child outcomes, which will be collected from participating children via assessments conducted by Elklan:

- Clinical Evaluation of Language Fundamentals Preschool-2 (CELF Preschool-2)
    - Basic Concepts
    - Concepts & Following Directions
- Early Years Toolbox: Early Numeracy Assessment

Elklan will administer the CELF Preschool-2 at the beginning of the study. At the end of the study, Elklan will administer the same test, along with the Early Numeracy Assessment.

We will also be asking you as a parent or guardian to participate in a short survey. We will do this at the start and end of the study.

A Data Sharing Agreement (DSA) will be put in place with your setting before any data is collected. The DSA will fully specify the data to be collected and for what purpose, who will collect it, how it will be shared, how it will be stored, who will have access to it and how long we will keep it for.

5. Who will hold your data?

During the period of the evaluation, no one outside the evaluation team will have access to your data. This project is one of several in the Stronger Practice Hub initiative, an aspect of the Department for Education (DfE) Early Years Educational Recovery Programme. As such, RAND Europe and the University of York will act as joint controllers for the data throughout the evaluation period. This means that both parties will share responsibility for ensuring your data and your child's data remain protected throughout the evaluation, including making sure that it is stored and shared securely. After the end of the evaluation, the DfE and EEF will become sole controllers of the data.

RAND Europe will lead on the processing of the pupil data – that is, they will be the main point of contact for any matters relating to the protection of all personal data and will make decisions about how and what personal data is used, in accordance with the purposes

agreed with the DfE and EEF. The University of York will lead on the processing of all staff data.

Better Communication CIC (the delivery team, funded by the EEF) as well as Elklan (commissioned by RAND Europe) will act as data processors throughout the evaluation period and will process the data, in accordance with the ways and purposes set by the joint controllers.

RAND Europe and the University of York are both registered with the Information Commissioner's Office, with registration numbers Z6947026 and Z4855807, respectively.

6. How will we collect, share and use your data?

As part of the evaluation, **Better Communication CIC** (the delivery team) will provide RAND Europe with recruited settings' contact details (i.e., email addresses, setting's postal address). **RAND Europe** and **Elklan** will use settings' contact details to collect information from settings about all the children that take part in Concept Cat (as specified in Section 4). **RAND Europe** will also ask Elklan to collect data on assessed child outcomes (as specified in Section 4). **Settings** and **Elklan** will provide this information by using an Excel data collection form provided by the evaluation team. This form will be shared via secure file transfer (e.g., Egress). The evaluation team will use the data provided by settings and Elklan on children participating in Concept Cat and compare their outcomes to similar pupils that did not receive the programme to measure if there is any difference between the two groups.

The **University of York** will ask Better Communication CIC (the delivery team) to collect information from settings about key staff. **Settings** will provide this information by using an Excel data collection form. This form will be shared via secure file transfer (e.g. Egress). **The University of York** will use the data provided by settings on key staff to invite them to take part in a short survey and/or interview. **RAND Europe** may also use this data in its research analysis to understand the impact of the programme.

At the end of the study, **RAND Europe** will submit the data in pseudo-anonymised format to the Office for National Statistics Secure Research Service (ONS SRS) for archiving in the **EEF** data archive (managed by EEF's archive manager).[14] This data will only be identifiable to the **DfE** and may be matched to the National Pupil Database (NPD) and other administrative data during subsequent research. The EEF and DfE will act as data controllers for the archive, along with contractors appointed to manage the archive.

[14] You can find more information about the EEF archive on the EEF's website: https://educationendowmentfoundation.org.uk/privacy-notices/privacy-notice-for-the-eef-data-archive.

No individual will be named in any report or other output for this project.

7. What is our legal basis for processing your data?

The legal basis for RAND Europe to process your child's personal data is legitimate interests, as detailed in Article 6(1)(f) of the UK GDPR. The legal basis for processing your child's special category data[15] is because it is necessary for archiving purposes in the public interest, scientific, or historical research purposes as detailed in Article 9(2)(j) respectively of the UK GDPR. To ensure that all processing is fair and lawful, RAND Europe have also completed a Legitimate Interest Assessment and a Data Protection Impact Assessment, in addition to the completion of an application to the RAND internal review board for ethical approval (see Section 8 below). RAND Europe will process only what is required to meet these legal bases and will ensure security and safeguards are in place to protect the information.

The legal basis for the University of York, EEF and DfE to process your child's personal data is where it is necessary for the performance of a task carried out in the public interest as set out in Article 6(1)(e) of the UK GDPR. The specific legislation which allows this is Section 10 of the Education Act 1996. The legal basis for processing special category data is for reasons of substantial public interest as detailed in Article 9(2)(g) of the UK GDPR.

8. Ethical Approval

The study has received ethical approval from RAND Europe's internal review board (HSPC ID: 2023-N0010, dated 14 March 2023).

9. How do we keep your data secure?

We take information security extremely seriously and we have appropriate technical and organisational measures to protect personal data and special category data. Access to information is restricted on a need-to-know basis and security arrangements are regularly reviewed to ensure their continued suitability. The evaluation team will collect and store all personal and special category data in accordance with the Data Protection Act (2018) and UK GDPR requirements.

No personal information collected as part of this study will be transferred outside of the European Economic Area (EEA).

For further information about how we will share your data, please refer to Section 6 above.

[15] 'Special category' data is personal data that needs more protection because it is sensitive, for example health or ethnicity data.

10. How long will we keep your data?

All individually identifiable data held by RAND Europe will be destroyed one year after the end of the study (2026). All individually identifiable data held by the University of York will be destroyed 5 years after the end of the study (2030). Data in the EEF's archive in the ONS SRS will include data only individually identifiable to the Department for Education (DfE), the government department responsible for children's services and education, and is kept indefinitely for the purposes of future research. Anonymous data will be kept indefinitely by the University of York.

11. What rights do you have in relation to your data?

Under the General Data Protection Regulation, you have rights in relation to yours and your child's data, including a right of access to the data (Article 15 of the GDPR), a right to rectification (Article 16 of the GDPR), erasure (in certain circumstances; (Article 17 of the GDPR)), restriction (Article 18 of the GDPR) and objection (Article 21 of the GDPR).

You are free to withdraw your child from the study at any time during the course of the study (information already collected about you will be retained). In this event, please tell your child's setting who will communicate this to the research team, or you can contact us directly using the details below.

If you wish to exercise any of the rights set out above in connection with this research project, please email us or write to us at the contact addresses below.

12. Right to complain

If you are unhappy with the way in which the research team has handled your personal data, we ask that you contact us in the first instance, to enable us to resolve your concerns. If you remain dissatisfied, you have the right to make a complaint to the Information Commissioner's Office (ICO), the UK supervisory authority for data protection issues, by emailing casework@ico.org.uk. Further information about how you can exercise your right to complain is available at www.ico.org.uk.

13. Changes to our privacy notice

We may change this Privacy Notice from time to time. If we make any significant changes in the way we treat your personal information, we will make this clear by contacting setting and ensuring they provide you with an updated version of this Privacy Notice.

14. How to contact us

If you have any questions about this privacy notice or concerns about how your data is being processed, you can get in touch by:

- Sending an email to the evaluation team FAO, Elena Rosa Brown (conceptcatevaluation@randeurope.org)
- Calling the evaluation team by ringing +44 1223 353 329 and quoting Ref: 022807.015 “Concept Cat”
- Contacting RAND Europe’s Data Protection Officer, Rani Viknaraja (rviknara@randeurope.org) and quoting Ref: 022807.015 “Concept Cat” or in writing to Data Protection Officer, RAND Europe, Westbrook Centre, Milton Road, Cambridge, CB4 1YG, UK
- Contacting University of York’s Data Protection Officer at dataprotection@york.ac.uk quoting Ref: “Concept Cat Evaluation”.

Appendix F: Information sheet and withdrawal form

CONCEPT CAT

PARENT/GUARDIAN INFORMATION SHEET

Dear Parent(s)/Carer(s),

Concept Cat is a new project that will be coming into your child's nursery. **Concept Cat** involves teaching children important words through songs, stories and play. The teaching is for all children and will be done by your child's usual teacher. We think that Concept Cat helps children learn, but we want to find out if it really does. To know we need to measure children's knowledge of words and maths. We are planning for a trained professional to come into nursery and talk to children in September 2023 and again in June/July 2024.

We gained full support from your child's nursery to deliver and evaluate Concept Cat, but if you don't want us to talk to your child to find out how they are learning words, you need to tell us. You can do this by filling in the 'I don't want my child to take part' form on page 6 or by emailing us on conceptcatevaluation@randeurope.org.

The full details of the **Concept Cat** project and its evaluation are below.

What is Concept Cat?

Concept Cat has been developed by Stephen Parsons and Anna Branagan, two Speech and Language Therapists, who are working with Better Communication CIC. Concept Cat is a story and play-based approach for teaching words that we call ‘concepts’. Concepts are really important words such as ‘first, whole and least’ which can be really tricky to learn. Concept Cat is a whole-class approach and so all children will be able to get involved. Each week a short story will be acted out and then the children will play some extra games and sing a song. Families can get involved when teachers send home the words and simple activities. Children will be taught early verbal concepts such as ‘first’, ‘wide’, ‘empty’ and ‘through’ in a structured and engaging manner that will involve storytelling and play. Such concepts have been selected as they are considered core to the curriculum of maths and science. The ultimate aim is to improve maths and science attainment at Key Stage 1. Children will be taught one word per week. This is because we want to allow opportunity for deep understanding to develop.

The evaluation

We want to find out if Concept Cat helps children learn words faster and if their numeracy skills develop so we will be running an evaluation of Concept Cat as it is delivered. The evaluation is being funded by the Education Endowment Foundation (EEF), an organisation that funds research into education, and the Department for Education (DfE). RAND Europe and the University of York with support from Elklan will be working together evaluate Concept Cat.

Settings that are selected to participate in the evaluation will be randomly assigned to either receive Concept Cat in the academic year 2023/2024 (the intervention group) or to continue practice as usual (the control group). There will be an equal chance of being allocated to each group, with half the settings in the intervention group and half in the control group.

The success of the programme will be measured using well-established language and numeracy assessments with selected children in the setting. In order to understand the impact of this approach, trained professionals from Elklan will visit nurseries at the start of the academic year and work with children one to one in September and October 2023 and then again in June/July 2024. They will talk to the children and look at some pictures for about 20-25 minutes each.

We are also keen to hear from parents and/or guardians so York University will also be asking you to participate in a short survey. They will do this twice: once at the

start and once again at the end of the school year. The purpose of the survey will be to understand your own experience of having your child participate in the programme, along with the types of learning activities that you and your child typically engage with at home (e.g., the simple activities that the teachers send home in order for the children to practise the new words).

What will taking part in the study involve?

Participation is voluntary

It is up to you if you want your child involved but you have to tell us if you don't want your child to take part.

If you choose not to include your child in the evaluation, this will have no impact on the usual care your child receives in their nursery, or on whether the nursery uses the Concept Cat programme or not. If you change your mind at any point during the study, you will be able to withdraw your child's participation without having to say why.

Anonymity and confidentiality

We won't tell other people about anything specific about your child.

We assure you that all information relating to you and your child will be treated in the strictest confidence and processed in accordance with General Data Protection Regulations (GDPR). Any sharing of the research findings (e.g. through reports, academic publications and presentations) will be in an anonymous format. **We will not mention your child's nursery or your child's name in any report or publication coming from the research**.

Storing and using your data

How we will look after the information we have about your child.

We will also ask your child's nursery to share some additional details about your child to help with our analysis.

All information about the nursery and your child will be stored securely by code number on a password protected computer. Any identifiable information will be stored separately from the data and will be destroyed at the end of the evaluation.

At the end of the evaluation your child's data will be shared with the Department for Education, the Education Endowment Foundation (EEF), FFT Education (EEF's data processor for the EEF data archive) in order to place it into the data into the National Pupil Database. This will allow your child's data to be used in a pseudonymised form (i.e. names are not shared) during subsequent research. We may also share anonymised data (i.e. data with no identifiable data that can be linked to your child) with other research teams.

To ensure that all processing is fair and lawful, the evaluation team have completed a Data Protection Impact Assessment and have sought and obtained ethical approval from an internal review board. Further details relating to GDPR, third parties, and confidentiality are provided in the Privacy Notice. Please read this information carefully. You will need to confirm that you have read and understood this information before proceeding.

Please note: If we gather information that raises concerns about your/your child's safety or the safety of others, or about other concerns as perceived by the researcher, the researcher may pass on this information to another person.

What do I need to do now?

If you are happy with all of this, you don't have to do anything. If you are happy for your child to take the short assessments and for your child's nursery to share your child's data with us, as described above, then you do not need to do anything.

If you aren't happy to get involved, then you need to tell us.

If you do NOT want to share your child's data as described above, please complete the attached 'Parent/guardian: I don't want my child to take part' form and hand it to your teacher who will then let us know.

You are free to withdraw your child's information for use in this evaluation at any time between now and August 2024. You can do that by either writing to the setting (who will then contact the evaluation team) or emailing the evaluation team directly at: conceptcatevaluation@randeurope.org.

Please note: Children whose parents have indicated that they wish to withdraw their child's data from the evaluation (i.e. the assessments) may still participate in the programme.

Questions or concerns

If you need to know anything else, please contact us.

If you have any questions about this participant information sheet or concerns about how your child's data is being processed, please feel free to contact the lead researcher, Elena Rosa Brown at conceptcatevaluation@randeurope.org. If you still have questions about the way your child's data will be processed, please contact our Data Protection Officer at REdpo@randeurope.org.

Thank you for taking the time to read this information.

CONCEPT CAT

PARENT/GUARDIAN: I don't want my child to take part

Only fill this form in if you do not want your child to take part in the assessments

If you are happy for your child to take part in the evaluation (i.e. the assessments) and for your setting to share your child's data with us, as described in the information sheet, then you do not need to complete this form.

If you do **NOT** want to your child to participate in the evaluation, please complete and return this form to your child's setting. Your child may still participate in the programme, even if you indicate that you wish to withdraw your child's data from the evaluation.

If you do not return a completed form, we will assume you are happy for your setting to share information for use in this evaluation and for your child to participate in the research.

☐ I **do NOT** want my child to participate in the research or for my child's data to be shared for the evaluation of the Concept Cat programme.

Child's Name (Please print clearly)…….....
.

Parent/Carer Name (Please print clearly)……....

Signature ……

Date…….

Setting Name ……

Name of setting staff member processing the withdrawal ……..

Hand in to nursery

Appendix G: Logic model mapped to implementation dimensions and research methods

Logic Model Outcomes	Aspect of the Logic Model	Data Collection											Implementation Dimensions							
		Monitoring data			Observations of Training Delivery	Setting Observations	Embedded Setting Observations	Practitioner Interviews	Baseline Practitioner Survey and quiz	Endline Practitioner survey and quiz and manager Survey	Baseline Parent Survey	Endline Parent Survey	Fidelity/Adherence	Dosage	Quality	Reach	Responsiveness	Programme Differentiation	Monitoring of Control Condition	Adaptation
	Moderators: P = Practitioner factors S = Setting factors H = Home factors	Coach visits to settings	Group supervision logs	Training Logs / attendance data																
Outputs	Practitioners engage in initial training support and materials to acquire understanding of programme approaches, including not to teach opposites simultaneously (P, S)																			
	Practitioner continuously deliver all elements of the approach (P, S)																			
	Practitioners engage in interim support sessions with peers and Word Aware trainers (P, S)																			

	Children experience increased explicit multi-sensory teaching of key conceptual vocabulary (explicit teaching) (P, S)																			
	Children have daily opportunities to hear and explore target concepts within play activities (implicit teaching) (P, S)																			
	In setting and at home, children have increased opportunities throughout the intervention period to review words that have been previously introduced (P, S, H)																			
	Families have increased opportunities to learn how to support their child's vocabulary																			
Short-term outcomes	Teachers gain understanding of the importance of conceptual vocabulary (P, S)																			
	Teachers have the skills and resources to effectively teach conceptual vocabulary: repetition, opportunities for embedding learning and implicit and explicit teaching																			
	Children show increased engagement with learning new words through participations in activities and adult/peer interactions																			
	Children show understanding and use conceptual vocabulary at school and at home																			

	Families demonstrate interest and engagement in pupil conceptual vocabulary development through regular attendance in the info sessions and watching videos																			
	Families engage in word learning activities at home using the provided guidance and suggested activities																			
Long-term outcomes	Quality teaching of early verbal concepts in daily practice																			
	Increased conceptual receptive vocabulary																			
	Increased early numeracy development																			

[i]