EVALUATION PROTOCOL

Nurturing Empathy before Transition (NEBT) a split-cohort clustered randomised controlled efficacy trial

Sheffield Hallam University

Principal investigator: Sarah Reaney-Wood

YOUTH ENDOWMENT FUND

Project title	*Nurturing Empathy before Transition (NEBT) a split-cohort clustered randomised controlled efficacy trial*
Developer (Institution)	*Roots of Empathy*
Evaluator (Institution)	*Sheffield Hallam University*
Principal investigator(s)	*Sarah Reaney-Wood*
Protocol author(s)	*Sarah Reaney-Wood, Bernadette Stiell, Claire Wolstenholme, Lucy Clague, Sean Demack, Jo Booth & Eleanor Byrne*
Trial design	*Split-cohort two-armed cluster randomised controlled trial with random allocation at the school level. Two Y5 pupil cohorts over 2 years (2022/23 & 2023/24).*
Trial type	Split Cohort (2 cohorts) Efficacy Trial
Evaluation setting	School
Target group	*Year 5 pupils*
Number of participants	*4200 pupils in 140 schools across two years (60 schools in year one and 80 schools in year two). One class of 30 Y5 pupils per school.*
Primary outcome and data source	***Updated**: Self-reported Me & My Feelings (M&MF) behavioural difficulties scale (Deighton et al, 2012)*
Secondary outcome and data source	***Updated:** Self-reported M&MF emotional difficulties scale (Deighton et al, 2012)* *Self-reported Basic Empathy Scale (BES) affective empathy and cognitive empathy scales (Joliffe & Farrington, 2006).* *Teacher reported Strengths & Difficulties Questionnaire (SDQ); total difficulties and prosocial scales (Goodman, 2001).*

	SDQ; hyperactivity, emotional problems, conduct problems and peer problems subscales (Goodman, 2001).

Protocol version history

Version	Date	Reason for revision
1.1 [*latest*]	March 2024	This updated protocol provides detail on the following six changes to the original protocol. These changes are explained in the first section. This section also highlights two additions for this protocol; updated versions of Appendix 1 and 2. 1. Changes to the evaluation team. 2. Change to primary outcome. 3. Correcting the number of training days for NEBT. 4. Highlighted the split-cohort design and geographical expansion to include Welsh schools in cohort 2. 5. Including a finalised NEBT logic model in the appendix. 6. Updated Impact/IPE activity undertaken for cohort 1 and 2 and further planned activity for cohort 2.
1.0 [*original*]	July 2023	*[leave blank for the original version]*

Table of contents

Protocol version history 2

Table of contents 3

Protocol Updates February 2024 5

Study rationale and background 8

Intervention 9

1. Named 9

2. Why 9

3. What (materials) 10

4. What (procedures) 10

Instructor Training: 10

NEBT delivery: 10

5. Who (provider) 11

6. How 11

7. Where 12

8. When and how much 12

Impact evaluation 12

Table 1: Trial design 14

Outcome measures 20

Longitudinal follow-ups 24

No longitudinal follow-ups will be undertaken as part of the NEBT RCT evaluation. However, Unique Pupil Identifiers (UPNs) are being collected by the evaluation team to enable long-term follow-up by others. 24

Implementation and process evaluation 25

Cost evaluation 34

Ethics and registration 35

Data protection 35

Stakeholders and interests 36

Risks 36

Timeline.....41

References 44

Appendix 1: Updated Impact and IPE activity and how this links to RQs and LM.....46

Appendix 2: Finalised NEBT logic model.....48

Appendix 3: EIF template logic model 49

Protocol Updates February 2024

1. Changes to the evaluation team.

The evaluations principal investigator (Sarah Reaney-Wood) is now on maternity leave. Additionally, the co-investigator (Bernie Stiell) will be leaving Sheffield Hallam University at the end of February 2024. Sean Demack has taken the principal investigator role while Sarah is away and is joined by Jo Booth and Eleanor Byrne. Sean will lead on the impact evaluation and will be joined by Giota Blouchou from March/April 2024. Sean is the deputy head of SIRKE and has extensive experience in RCTs involving young people. Giota has a background in developmental psychology and will join SIRKE in March 2024 as a Research Fellow in Quantitative Education Research. Giota will shadow Sean in the hand-over of aspects of the impact evaluation from data collection through analyses and into write up. Jo and Eleanor will undertake the remaining IPE activities following the final hand over from Bernie at the end of February 2024. Both Jo and Eleanor have extensive IPE experience.

The extent of change to the evaluation team has been notable; a total of eight SHU academics have been involved during the five years of the project with two of these people (Sean Demack and Sarah Reaney-Wood) having consistent involvement. Some of this can be accounted for by the timescale of the project and the impact of Covid 19. SHU were appointed as evaluators for the first round of YEF projects 2019 but the evaluation was postponed because of the Covid 19 pandemic so that recruitment took place in the 2021/22 academic year. The post-pandemic difficulties in recruitment led to the decision to adopt a split-cohort design with cohort 1 running in 2022/23 and cohort 2 running in 2023/24. This meant that the initial plans for a two-year project and one-year evaluation in 2020/21 (recruitment in 2019/20) expanded to a five year project with two one-year evaluation cohorts; cohort 1 in 2022/23 (recruitment in 2021/22) and cohort 2 in 2023/24 (recruitment in 2022/23).

2. Change to the primary outcome

The original protocol identified the total score from the Me & My Feelings questionnaire as the primary outcome for the impact evaluation. The peer reviewer for the Statistical Analysis Plan (SAP) for the evaluation highlighted that the M&MF questionnaire is not psychometrically designed as an overall scale, but as two subscales (behavioural difficulties and emotional difficulties). The primary outcome is now specified as the M&MF behavioural difficulties scale. The M&MF emotional difficulties scale has been included as an additional secondary outcome. This change will result in altering the Research Questions (RQs) for the impact evaluation which can be found in the Statistical Analysis Plan (SAP).

3. Correcting the number of face-to-face training days for NEBT

The original protocol stated that instructors would receive two days of initial training when in reality there are four: three initial training days plus one mid-year training day. This was the case for both cohorts 1 (2022/23) and 2 (2023/24). This correction results in re-defining the training attendance criteria for fidelity and compliance analyses which can be found detailed in the SAP.

4. Highlighting the split-cohort design and geographical expansion to include Welsh schools in cohort 2.

The original protocol stated that "The intervention will run for two consecutive years, for around 9 months of the year, involving a total of 140 schools (70 intervention and 70 control). First, from Autumn 2022 to Summer 2023, involving 60 schools (30 intervention and 30 control) and second, from Autumn 2023 to Summer 2024 involving 80 schools (40 intervention and 40 control)."[1] However, we feel that the split-cohort aspect of the RCT design should be more clearly stated; and this is why it is now included in the title of the redrafted protocol. Split-cohort designs pool data for multiple cohorts (Y5 pupils in two consecutive years in this instance) for gains in statistical sensitivity (or power). Combining data from multiple RCT-centred evaluations can be useful when delivery capacity is limited and/or if recruitment is lower than expected but does assume that it is reasonable to do. Assessing this assumption draws on IPE and impact evaluation analyses that examine consistency in terms of impact, delivery, fidelity and compliance across the cohorts.

Please see the SAP for a CONSORT flow diagram for the primary outcome (M&MF behaviour difficulties scale) drawing on baseline and outcome data for cohort 1 and baseline data for cohort 2. In summary, for cohort 1 58 schools signed a Memorandum of Understanding (MoU) by summer 2022, only 33 of these completed the baseline data collection requirements to be randomised in October 2022. This led to a need for a larger sample for cohort 2 and, to help achieve this, the geographical reach of the programme was extended to include Wales. For cohort 2, 106 schools signed a MoU by summer 2023 with 54 of these completing the baseline data collection requirements to be randomised in October 2023.

[1] See p10 of original protocol; https://youthendowmentfund.org.uk/funding/who-we-fund/roots-of-empathy/

The geographical expansion along with time/cohort are two known systematic differences between the evaluations in cohorts 1 and 2. This highlights the need for additional sensitivity analyses for this split-cohort efficacy RCT design; and these are detailed in the SAP.

5. Including a finalised NEBT logic model in the appendix.

The original protocol provided a draft NEBT Logic Model as Appendix 2. The NEBT logic model was finalised in July 2023 and has been included as Appendix 2a (following the original Appendix 2).

6. Updated Impact/IPE activity undertaken for cohort 1 and 2 and further planned activity for cohort 2

In Appendix 1 of the original protocol, a Table linked Impact and IPE activity to specific research questions and Logic Model codes. This has been updated to link to the finalised Logic Model and included as Appendix 1a (following the original Appendix 1).

Additionally, IPE activity assumed around 60 schools (30 in the NEBT group) in cohort 1 and 80 schools (40 NEBT) in cohort 2 (140 schools in total). The achieved samples were lower with 33 schools (16 NEBT) in cohort 1 that dropped to 19 schools with baseline & outcome data for the primary outcome. In cohort 2, there were 54 schools randomised at baseline. Across both cohorts there were 87 schools randomised at baseline with a maximum of 73 schools with baseline & outcome data for the primary outcome. The smaller sample resulted in reduced IPE data collection activities for both cohorts 1 and 2 and this is shown in Appendix 1a.

Study rationale and background

There is extensive evidence for the benefits of and need for well-designed school-based interventions that focus on developing pupils social and emotional wellbeing (Browne et al., 2004; Durlak et al., 2011; Tome et al., 2021). School based interventions have been demonstrated to lead to improvements in behaviour and learning (Panayiotou et al., 2019); academic success, better health outcomes and later life success.

This need for well-designed and implemented interventions is thought to have increased since the COVID-19 pandemic (Lee et al., 2020; Hamoda et al., 2021) as the Covid-19 pandemic has negatively impacted child and youth mental health and wellbeing (Office for Health Improvement & Disparities, 2022).

Youth violence has been increasing around the world in recent years and is resulting in serious societal costs (Haylock et al., 2020). The United Kingdom has consistently seen an increase in the incidence of youth violence since 2012/13 (Haylock et al., 2020). Specifically, poor mental health has been associated with violence among youth, including gang violence. During a time when the Covid-19 pandemic has negatively impacted child and youth mental health and wellbeing (Office for Health Improvement & Disparities, 2022), programmes that mitigate this negative impact and that support positive mental health are increasingly important.

Current evaluation in the context of previous evaluations of Roots of Empathy programme

There have been several previous evaluations of the Roots of Empathy programme with the majority employing a quasi-experimental design with two employing a Randomised Control Trial (RCT) design. Previous evaluations of the Roots of Empathy programme have demonstrated that the programme may lead to an immediate increase in prosocial behaviour (teacher perception), reduction in problem behaviour and an increase in understanding of infant development (Connolly et al, 2018). The evaluation by Connolly et al was conducted in Northern Ireland with pupils aged 8-9, where this research is in England with students aged 9-10 years. There has been other evaluations of the RoE programme that have focused on a range of age groups.

This evaluation builds on previous evaluations, including developing on previous methodological weaknesses (e.g. an over-reliance on teacher reports) as well as expanding on the geographical locations where robust evaluations of NEBT have taken place. The causal impact of NEBT on the social and emotional development of year five pupils will be estimated using two clustered RCT designs. Participating primary schools will be drawn from throughout England.

Intervention

Tidier framework

1. Named

Roots of Empathy Nurturing Empathy Before Transition (RoE NEBT)

2. Why

Roots of Empathy (RoE) have designed the Nurturing Empathy Before Transition (NEBT) programme, which aims to increase empathy and prosocial behaviour in school children in Year 5. This well-established programme involves bringing a parent and baby into the classroom as part of a structured programme of lessons focused on building empathy. It is described by the delivery partners as "an evidence-based, preventative intervention for primary school children, that aims to reduce aggression, including bullying, and increases children's social and emotional competence."

The programme is underpinned by the assumption that empathy is innate and the extent to which it develops is dependent upon the attachment relationships children build. The RoE programme teaches both the cognitive and emotional elements of empathy by encouraging pupils to identify the baby's feelings, whilst they also reflect on their own feelings and the feelings of others, and thus improving their emotional literacy. This improved emotional literacy alongside witnessing the mother regulating the baby's emotions, enables children to better regulate their own emotions, leading to improvements in emotional regulation, resilience and wellbeing. This in turn leads to reduced aggression and an increase in prosocial behaviour. Please see logic models in appendices two and three.

In line with the above theoretical assumptions, a previous randomised controlled trial of the RoE programme in Northern Ireland found that the programme was well received in schools and that a positive effect in teacher rated prosocial behaviour could be observed (Connoly et al, 2018). Furthermore, previous studies (mainly quasi-experimental (QED) have highlighted that RoE may lead to decreased aggression (e.g. Santos et al., 2011; Latsch, Stauffer and Bollinger, 2017) and increased empathy (Wrigley, Makara & Elliot, 2015; Latsch, Stauffer & Collinger, 2017). However, an evaluation of the programme and its effects in England has not yet been conducted.

3. What (materials)

NEBT instructors are provided with teacher guides and supporting materials which should enable them to teach the 27-session programme of NEBT in one academic school year.

4. What (procedures)

Instructor Training:

ROE will train instructors from within participating schools e.g. teaching assistants. Instructors will receive specialist training in how to deliver the NEBT programme through 4 days of face-to-face training (3 initial days and 1 mid-year day).

NEBT delivery:

For cohort 1, The NEBT intervention took place within four geographical regions (Yorkshire, Merseyside, East & West Midlands & Greater London)[2]. To assist with meeting recruitment targets, this was expanded to five areas with schools in Wales included.

Instructors then deliver the NEBT in Y5 classes in intervention schools. The programme consists of 27 sessions, split into 9 themes, with 3 sessions per month. The parent and baby attend one session per month, with a preparation and debrief session either side. The 9 broad themes which the sessions are based around are listed below:

Theme 1 – Meeting the Baby

Theme 2 – Crying

Theme 3 – Caring and Planning for Baby

[2] Recruitment originally focused on five smaller geographical areas (Doncaster, Birmingham, Northamptonshire, Nottingham and London) but this was changed to four larger areas (Yorkshire, Merseyside, the Midlands and Greater London) following initial recruitment difficulties.

Theme 4 – Relationships

Theme 5 – Sleep

Theme 6 – Safety

Theme 7 – Communicating

Theme 8 – Who am I?

Theme 9 – Goodbye and Good Wishes

Sessions last for approximately 40-45 minutes. However, as the welfare of the mother and baby are of paramount importance it is feasible that the mother and baby session may be shorter if needed.

5. Who (provider)

Roots of Empathy[3] is a children's charity whose mission is to build caring, peaceful and civil societies through the development of empathy in children and adults. ROE Instructors will be members of staff based in participating schools, most likely a teaching assistant or SENCO who will be trained prior to the start of the intervention by Roots of Empathy. The class teacher will be present during sessions to observe and deal with any behaviour management as needed.

6. How

The NEBT intervention is delivered face-to-face in the school setting and is delivered to the whole class. Sessions take place when the schools feel it is appropriate within the timetable, but usually replaces a Personal, Social, Health and Economic (PSHE) session, as ROE feel the programme contributes to many of the same learning objectives. Each Roots of Empathy visit has a specific lesson plan that the instructor will follow step by step. The family visit focuses on guided observation, discussion and interaction with the parent and baby and as such will

[3] See https://rootsofempathy.org/

be different each time. The pre and post visits include questions and discussion, group work, art and storytelling.

7. Where

NEBT will run in four geographical locations in the UK: Yorkshire, Merseyside, East & West Midlands & Greater London. 70 schools located within these areas will receive the intervention. Schools will be recruited from areas of social disadvantage, which ROE have classified as schools with over 21% pupil premium/Free School Meals in any of the 4 broader geographical areas.

8. When and how much

The intervention runs for one academic year but involved two Y5 pupil cohorts; cohort 1 in 2022/23 and cohort 2 in 2023/24. In both cohorts, the intervention will run for around 9 months of the year, involving a total of 140 schools for the combined cohorts (70 intervention and 70 control). First, from Autumn 2022 to Summer 2023, involving 60 schools (30 intervention and 30 control) and second, from Autumn 2023 to Summer 2024 involving 80 schools (40 intervention and 40 control). In each intervention school, one year 5 class (of approximately 30 pupils) receives the intervention. Three sessions will be delivered to the class each month.

Impact evaluation

Research questions or study objectives

Primary research question

1. RQ1: (Impact evaluation-primary outcome) What is the impact of the Nurturing Empathy programme on self-reported behavioural difficulties of primary school aged children when compared to a 'business as usual' control?

Secondary research questions

2. RQ2: What is the impact of the Nurturing Empathy programme on self-reported emotional difficulties of primary school aged children when compared to a 'business as usual' control?

3. RQ3: What is the impact of the Nurturing Empathy programme on self-reported affective empathy of primary school aged children when compared to a 'business as usual' control? (secondary outcome)

4. RQ4: What is the impact of the Nurturing Empathy programme on self-reported cognitive empathy of primary school aged children when compared to a 'business as usual' control? (secondary outcome)

5. RQ5: What is the impact of the Nurturing Empathy programme on teacher-reported prosocial behaviour of primary school aged children when compared to a 'business as usual' control? (secondary outcome)

6. RQ6: What is the impact of the Nurturing Empathy programme on teacher-reported school behaviour of primary school aged children when compared to a 'business as usual' control? (secondary outcome)

Exploratory

7. What is the difference in teacher-reported emotional problems (SDQ Sub scale) between the intervention group, when compared to a 'business as usual' control?

8. What is the difference in conduct problems (SDQ Sub scale) between the intervention group, when compared to a 'business as usual' control?

9. What is the difference in peer relationship problems (SDQ Sub scale) between the intervention group, when compared to a 'business as usual' control?

10. What is the difference in hyperactivity (SDQ Sub scale) between the intervention group, when compared to a 'business as usual' control?

Design

This efficacy trial has a 2-armed, two-level RCT design where pupils are clustered within school. The unit of randomisation is the school. The allocation is 50:50, with the aim for an equal number of schools allocated to intervention and control. Randomisation will be stratified by geographical area to ensure similar numbers of intervention and control schools in each of the four geographical areas.

The primary outcome is the Me and My Feeling questionnaire Behavioural Difficulties subscale[4] and the secondary outcomes are the Me and My Feeling questionnaire Emotional Difficulties subscale, the Basic Empathy Scale (BES) Affective and Cognitive Empathy subscales[5] the Strengths and Difficulties questionnaire (SDQ)[6] total difficulties score, prosocial, emotional problems, conduct problems, hyperactivity and peer problems subscales. More information on what these scales measure and how they will be analysed is given below. Data for the M&MF and the BES will be collected directly from the pupils, online. The data on the SDQ will be collected indirectly from the class-teacher, online.

Table 1: Trial design

Trial design, including number of arms		*Split-cohort, two-armed, cluster randomised controlled trial*
Unit of randomisation		*School*
Stratification variables (if applicable)		*Geographic area*
Primary outcome	variable	*Behavioural difficulties at outcome*
	measure (instrument, scale, source)	*Behavioural difficulties as measured using the Self-report Me and My Feelings questionnaire behavioural difficulties sub scale (Deighton et al, 2013) [0 to 12 scale]*
	variable(s)	*Emotional difficulties, cognitive empathy, affective empathy, teacher reported behaviour difficulties, prosocial behaviour,*

[4] See https://www.corc.uk.net/outcome-experience-measures/me-and-my-feelings-mmf/

[5] See https://www.rand.org/education-and-labor/projects/assessments/tool/2006/basic-empathy-scale-bes.html

[6] See https://www.sdqinfo.org/py/sdqinfo/b3.py?language=Englishqz(UK)

Secondary outcome(s)		*hyperactivity, emotional problems, conduct problems and peer problems*
	measure(s) (instrument, scale, source)	*M&MF Emotional Difficulties subscale, [0 to 20 scale]* *BES cognitive empathy subscale, [9 to 45 scale]* *BES affective empathy subscale.[11 to 55 scale]* *Teacher SDQ total difficulties subscale [0 to 40 scale]* *Teacher SDQ prosocial score [0 to 10 scale]* *Teacher SDQ hyperactivity score [0 to 10 scale]* *Teacher SDQ emotional problems score [0 to 10 scale]* *Teacher SDQ conduct problems score [0 to 10 scale]* *Teacher SDQ peer problems score [0 to 10 scale]*
Baseline for primary outcome	variable	*Behavioural difficulties at baseline*
	measure (instrument, scale, source)	*M&MF Behavioural difficulties subscale [0 to 12 scale]*
Baseline for secondary outcome	variable	*Emotional difficulties, cognitive empathy, affective empathy, teacher reported behaviour difficulties & prosocial behaviour*
	measure (instrument, scale, source)	*M&MF Emotional Difficulties subscale, [0 to 20 scale]* *BES cognitive empathy subscale, [9 to 45 scale]* *BES affective empathy subscale.[11 to 55 scale]* *Teacher SDQ total difficulties subscale [0 to 40 scale]* *Teacher SDQ prosocial score [0 to 10 scale]* *Teacher SDQ hyperactivity score [0 to 10 scale]* *Teacher SDQ emotional problems score [0 to 10 scale]* *Teacher SDQ conduct problems score [0 to 10 scale]* *Teacher SDQ peer problems score [0 to 10 scale]*

Randomisation

Randomisation will be undertaken by the evaluation team at Sheffield Hallam University (SHU). All year one schools will be randomised in September 2022, following baseline testing at the start of year 5. Randomisation will take place at the school level, with schools organised into four geographical locations. Simple randomisation will take place within each geographical location. SHU will communicate the results of randomisation to NEBT intervention and control schools.

Randomisation will be conducted in Microsoft excel as follows: schools will be organised into groups by geographical location. Within each geographical group, schools will then be assigned a value through a random number generator, schools will then be sorted by the random number (within group), and treatment or control allocation will be assigned using the ABABAB pattern within each group. Schools will be informed of their allocation by SHU .

Control schools will be passive, meaning they will not undertake any activities as part of this trial. Schools will continue their business as usual (BAU), a normal school lesson as part of their school curriculum. Control schools will receive an incentive payment of £400 to recognise the time taken by teachers and pupils to complete the baseline and endpoint data collection.

Participants

Recruitment to the NEBT trial is being managed by Roots of Empathy and is being conducted at the school level (one class of 30 pupils in each NEBT intervention school). Where schools have more than 1-form entry, schools are being given the option of which class to include. This follows RoE usual protocol for schools with more than 1-form entry, the same process is being undertaken for control and intervention schools. As the evaluation of NEBT trial is taking place across two years (two separate efficacy trials), ROE school recruitment will take place twice.. For schools to be eligible to take part in the NEBT trial they must meet the following inclusion and exclusion criteria:

Inclusion criteria:

- Be a primary school in the following geographical areas; Yorkshire, Merseyside, the Midlands (East and West) and Greater London (cohort 1) expanded to include Wales in cohort 2[7]
- Have >21% FSM or PP

Exclusion criteria:

- Prior experience with a Roots of Empathy programme
- Private schools, special schools or PRU

[7] The original plan was to focus on five smaller geographical areas (Doncaster, Birmingham, Northamptonshire, Nottingham and London) but RoE struggled to recruit in these areas and so the geographical scope was widened to include the stated four larger geographical areas.

The NEBT intervention will be delivered to year 5 pupils, in school, in a classroom allocated to the recipient class. Where schools are more than one form entry, the school will decide which Y5 class will receive the NEBT intervention. Data collection for the evaluation will be undertaken in schools, within the pupil's usual classroom or computer room. Guidance to facilitate data collection will be provided by SHU to ensure consistency between schools. The questionnaires (M&MF, BES and SDQ) will be conducted online as a preference. Schools will be given the opportunity for paper-based copies on a school-by-school basis, if it is felt this would be beneficial in facilitating both trial involvement and data collection.

Sample size calculations

Sample size was determined prior to recruitment taking place by calculating minimal detectable effect sizes (MDES) undertaken by the evaluation team. Specifically, the MDES is the estimated smallest difference (between the intervention and control groups) in the primary outcome that the design could detect as being statistically significant ($p<0.05$, two tailed) with a statistical power of 0.80 or greater. This 'smallest differences' is presented as a standardised Hedges g effect size in units of standard deviations.

The power analyses drew on data available from a previous RCT (Connolley et al, 2018) reported effect sizes of +0.20; p = 0.05 for prosocial behaviour, meaning the intervention group were rated as more prosocial by their teachers and -0.16; p=-0.06 for difficult behaviour, meaning the intervention group exhibited less difficult behaviour than those in the control group. We therefore looked to design an evaluation with enough sensitivity to detect similar effect sizes. Discussions were then had between evaluator, delivery and funder teams to best ensure that the MDES calculations were based on feasible estimates in terms of practical constraints and delivery capacity for NEBT. Given the level of evidence already available to the impact of the RoE programme, it was felt that it was most appropriate to ensure that the trial was appropriately powered from the start as an efficacy trial. A 2-level CRT design that could detect an MDES of 0.20 sds or higher (effect size found in last RCT), would need a total of 140 primary schools (with one class of 30 pupils per school), with the NEBT intervention running in 70 of these primary schools. This presented notable capacity challenges for RoE and a pragmatic solution was identified. Spreading the NEBT delivery across two years and treating these as both two distinct (but underpowered), and one combined (fully powered) efficacy trial provided the necessary flexibility in the scale of NEBT delivery, without compromising on the overall statistical sensitivity. Our two-trial design means that NEBT would be delivered in 30 schools for year one (from Sept 2022) and 40 schools in year two (from Sept 2023). Data from both years will be combined to maximise the sample size (and hence statistical sensitivity). For this design to be valid, it is important that the same NEBT programme is delivered in both years (i.e. the ToC should remain

consistent) and that additional external factors do not result in making the two years very 'distinct' (e.g. if one year experienced notable school disruption due to the re-emergence of a pandemic). Both impact evaluation and IPE will examine this. Within the impact evaluation, the impact in year 1 and 2 will be shown separately and then combined into the 'year 1+2' impact estimate. Within the IPE, data will be collected that will enable the implementation of NEBT in years 1 and 2 to be compared with reference to the underlying ToC and the context of delivery.

As explained above, the impact evaluation of NEBT will be undertaken using two efficacy trials that both have a 2-level clustered RCT design. The first trial will begin in Autumn 2022 and the second trial will begin in Autumn 2023, both trials lasting for around nine months. Data from both efficacy trials will be drawn together to maximise the sample size and statistical power, but the impact findings for each individual trial will be reported. Therefore, we present three MDES estimates for the first, second and combined efficacy trials respectively:

1. MDES for year one, 60 schools, 1800 pupils
2. MDES for year two, 80 schools and 2400 pupils
3. MDES for both years together 140 schools and 4200 pupils

Estimations for pre-post test correlations of the M&MF questionnaire are limited. As such, we have used a conservative estimate of 0.50. This is taken from estimates achieved for similar measures (for example, SDQ). The intra-cluster correlation is also estimated to be between 0.1-0.2. An updated version of power calculations will be made after year one of the trial has ended, drawing on the empirical data from the first efficacy trial.

Calculations were conducted in Excel using the formula from Bloom et al (2007)[8] and checked using the Powerup! software (Dong, 2015)[9]. Calculations for a 2-level RCT, with schools 60 per comparison and 30 pupils per school have been carried out. The target number of schools for recruitment over the two years is 140. The results of this analysis estimate that, for the main ITT analyses combining years 1 and 2, the design will be able to detect an effect size of between 0.16 and 0.22 sds or higher as statistically significant (α<0.050) with a statistical power of 0.80.

[8] Bloom, H.S., Richburg-Hayes, L. and Black, A.R. (2007) Educational Evaluation and Policy Analysis, Vol. 29, No. 1, pp. 30–59

[9] Dong, N., Kelcey, B., Maynard, R. and Spy brook, J. (2015) PowerUp! Tool for power analysis

Table 2: Sample size calculations

		Year 1 estimates	Year 2 estimates	Year 1 & 2 combined
Minimum Detectable Effect Size (MDES)		0.25-0.34	0.22-0.29	0.16-0.22
0. Pre-test/ post-test correlations	level 1 (participant)	0.50	0.50	0.50
	level 2 (cluster)	0.25	0.25	0.25
Intracluster correlations (ICCs)	level 2 (cluster)	LOW=0.1, HIGH=0.2	LOW=0.1, HIGH=0.2	LOW=0.1, HIGH=0.2
Alpha[10]		0.05	0.05	0.05
Power		0.8	0.8	0.8
One-sided or two-sided?		Two-sided	Two-sided	Two-sided
Number of clusters[11]	Intervention	30 schools	40	70
	Control	30	40	70
	Total	60	80	140
Number of participants pupils per school	Intervention	900 pupils	1200	2100
	Control	900	1200	2100

[11] Please state how the data is clustered, if there is any clustering (e.g. by delivery practitioner or setting).

		Year 1 estimates	Year 2 estimates	Year 1 & 2 combined
(working on average class size of 30)	Total	1800	2400	4200

Outcome measures

Baseline measures

Primary outcome (self-reported behavioural difficulties)

The primary outcome for the NEBT trial is behavioural difficulties as measured using the Me and My Feeling (M&MF) questionnaire (Deighton et al., 2013). M&MF is a 16-item school-based measure of child mental health, suitable for children as young as 8 years of age, covering two domains, emotional difficulties, and behavioural difficulties (Deighton et al., 2013). The behavioural difficulties score for M&MF will be used as the primary outcome, with the emotional difficulties subscale as secondary outcomes.

The NEBT programme aims to tackle the root causes of challenging behaviours by increasing social and emotional wellbeing. The evidence based short and medium-term outcomes detailed within the NEBT ToC relate to an increase in emotional literacy, and a decrease in violent and aggressive behaviours. As such, the M&MF questionnaire is the most age - appropriate way to measure some of these outcomes.

The psychometric properties of the M&MF are broadly good. This is demonstrated by good internal consistency (Deighton et al., 2013; Patalay et al., 2014), construct validity (Deighton et al., 2013), convergent validity with the subscales of the SDQ (Deighton et al., 2013) and good discriminant validity (Deighton et al., 2013).

Data for the primary outcome will be collected at baseline, pre-randomisation (September/October 2022 for year one schools and September/October 2023 for year two schools) and at endpoint (June/July 2023 for year one schools and June/July 2024 for year two schools.

M&MF is publicly available at Me and My Feelings (M&MF) (corc.uk.net)

Secondary outcomes

A total of nine secondary outcomes will be used; self-reported emotional difficulties, self-reported affective and cognitive empathy and six teacher-reported measures drawn from the SDQ (total difficulties, prosocial score, emotional problems, conduct problems, peer problems and hyperactivity).

Empathy

Empathy will be assessed at baseline and endpoint using the Basic Empathy Scale (BES) (Joliffe et al., 2006). The BES is a pre-validated scale to assess empathy in young people aged 9-18, focusing on cognitive and affective empathy. Whilst other existing measures focus on the three elements of empathy (rather than two as with the BES), the age group of the young people in the NEBT trial makes the BES the most appropriate.

As per the ToC, the RoE programme aims to improve empathy, prosocial behaviour and wellbeing. As such, the BES is an age-appropriate tool to assess this, using pupil self-report as opposed to parent or teacher perceptions.

Exploratory factor analysis of the BES demonstrated a two-factor structure, cognitive empathy and affective empathy, and this was confirmed using confirmatory factor analysis (CFA). The BES is a 20-item questionnaire with acceptable internal consistency .77-87. Confirmatory factor analysis (CFA) showed adequate model fit and test re-test were between r=.54-r=.70.

Strengths and difficulties questionnaire-behaviour

Pupil behaviour and mental health will also be assessed using the teacher version of the Strengths and Difficulties questionnaire (SDQ) (Goodman, 2001). The SDQ, teacher version is a 25-item scale used to assess behaviour in the school context in 4–16-year-olds.

The SDQ is commonly used in clinical assessments and has become increasingly popular as an outcome measure in a variety of evaluations. The SDQ has good psychometric properties with good internal consistency, (Cronbach a .62-.85) and test-retest reliability of r=.62. In addition, the SDQ is a multi-respondent measure with inter-rater reliability of .37-.58 for self-report and parent-report and .24-.39 for self-report and teacher report.

The teacher who completes the SDQ will need to be familiar with the pupils they are completing the SDQ for, as they are being asked to assess that pupil's regular behaviour. Additionally, it will be important that the same teacher completes to SDQ at baseline and outcome, where at all possible. Specific guidance on how much time a teacher needs to spend with a pupil prior to completing the SDQ is limited, but some materials from a London

school[12] states that "The SDQ needs to be completed by someone that knows the child or young person well and has regular contact with them. Regular contact can be viewed as someone that has had involvement with the child or young person 3 or 4 times a week for at least 3 months but preferably 6 months or more." Further, this guidance states "It is crucial that the person completing the SDQ is not basing their answers on a specific day but rather over a period of time. Otherwise, the results may not provide an accurate reflection of the child or young person's needs." In the context of the RoE NEBT intervention, primary school teachers are likely to have daily contact with their Y5 class of at least 5 hours a day. The baseline data collection will commence in week 4 of the Autumn term (in 2022 and 2023 respectively). The four weeks preceding this will be used to provide participating pupils (and parents) with information sheets and for schools to gather consent if this is what they decide. Additionally, by the start of week 4, we assume that primary teachers will have spent sufficient time with their pupils (at least 75 hours) to validly complete an SDQ for each of them.

Compliance

Compliance will be measured at the school, teacher/TA and the individual pupil level, and will be related to the number of sessions delivered and the number of sessions that the YP has attended.

School level compliance

School level compliance will be measured by the number of NEBT sessions delivered across the 9-month period. Data collected will be used to create a binary variable that indicates school level compliance, or not. The minimum threshold for number of NEBT sessions delivered in a school will be agreed and published in the Statistical Analysis Plan.

Instructor (Teacher/TA) level compliance

Teacher/TA level compliance will be measured by the number of RoE NEBT instructor training sessions that a teacher/TA attends. This will be measured on a categorical scale from 0 to 2.

Pupil level compliance

Registers will be taken at each session (x3 each month) and a YP will be considered to have complied if they have attended sessions covering 8 out of 9 topics covered by RoE.

[12] See https://www.afcvirtualschool.org.uk/attachments/download.asp?file=117&type=pdf

Data collected will be used to create a binary variable that indicates pupil, TA/teacher and school level compliance, or not. The minimum threshold for the number of non-child and child sessions that a pupil needs to attend to be classed as 'compliant' will be agreed and published in the Statistical Analysis Plan.

The final compliance variable will draw together the school- teacher/TA- and pupil-level compliance variables into a single binary pupil-level variable. The precise definition of the final compliance variable will be determined by the thresholds agreed for the three levels. For example, the final variable might be a pupil who attended X% of sessions located in a class taught by a teacher who attended both NEBT training sessions in a school that covered at least 8 out of the 9 NEBT topics .

Analysis

The NEBT RCT is taking place across two years, with 60 schools in the first year and 80 in the second year the headline/primary analysis will combine data from the two cohorts with sensitivity analyses examining impact with the two separate cohorts.

The primary outcome measure for this trial is the Me and My Feeling (M&MF) behavioural difficulties subscale taken at endpoint, at the end of Y5. An intention to treat (ITT) approach will be taken.

To answer RQ1[13], multi-level linear regression models will be constructed that acknowledge that pupils are clustered in schools. In each of these two models, the endpoint M&MF score will be the outcome variable with the trial arm (1=NEBT or 0=Control) as the independent variable and baseline M&MF and geographical location as covariates.

The impact of NEBT will be estimated by converting the model coefficient for the trial arm variable into Hedges' g effect sizes using the equation below, where T is the treatment mean, C is the control mean, δ^2_{sch} is the school level variance and δ^2_{pup} is the pupil level variance:

$$ES = \frac{(T - C)_{adjusted}}{\sqrt{\delta^2_{sch} + \delta^2_{pup}}}$$

[13] RQ1: What is the impact of the Nurturing Empathy programme on self-reported behavioural difficulties of primary school aged children when compared to a 'business as usual' control?

For the primary outcome analysis and follow-on exploratory analyses, statistical uncertainty will be expressed as standard errors of multilevel model coefficients and use of 95% confidence intervals.

Secondary outcome analyses

The secondary analysis to answer RQs 2 to 6 will each employ a multi-level linear regression model with pupils clustered in schools. For each model, the relevant secondary outcome measure (as listed above pp.8) will be the outcome/dependent variable and the explanatory variables will comprise the trial arm (1=NEBT or 0=Control), relevant secondary baseline measure (as listed above pp.8) and geographical location. As with the analyses of the primary outcome, the impact of NEBT will be estimated by converting the model coefficient for the trial arm variable into Hedges' g effect size.

Exploratory analysis

To answer the exploratory RQs 7-10, the same approach used for RQs 1-6 will be conducted but focusing on one of the four SDQ total difficulties subscales (included as the outcome and a baseline covariate).

Missing Data

The baseline and ITT samples will be compared to help illustrate the impact of missing data for the primary outcome variable only, Me and My Feelings (M&MF). This will firstly be done descriptively by tabulating missing cases across the categories of variables included in the ITT analysis. Reasons for any missingness will be summarised and we will examine whether missingness is associated with school and/or pupil-level covariates for example; baseline M&MF. Further detail on missing data analysis will be provided in the Statistical Analysis Plan (SAP).

Longitudinal follow-ups

No longitudinal follow-ups will be undertaken as part of the NEBT RCT evaluation. However, Unique Pupil Identifiers (UPNs) are being collected by the evaluation team to enable long-term follow-up by others.

Implementation and process evaluation

Research questions

a) What are the key factors which influence successful delivery of the NEBT programme in years 1 and 2? (PM interview, and case study visits in year 1 &2)
b) What are the perceptions of pupils, teachers, deliverers and instructors about the effectiveness of the programme in years 1 and 2? (Case study visits)
c) What fidelity issues are observed during years 1 and 2 of the trial?
d) What does the trial indicate about scalability?

In order to answer these research questions, the evaluators worked with the ROE team to develop an initial evidence-based logic model (see appendix 2). This is grounded in the existing research evidence on empathy and pro-social interventions, outlining the inputs, mechanisms and intended short, medium and longer-term outcomes of the NEBT programme. Contextual or moderating factors identified as likely to impact the delivery, fidelity and outcomes of the NEBT programme will also be explored over the two years of the evaluation, informed by the logic model. It is assumed that the intervention in years 1 and 2 are comparable, so the way the logic model is implemented and the intervention delivered in both years is assumed to be the same.

When the quantitative and qualitative evidence from the two years is combined and compared, it will be used to test and update the logic model, identifying other possible causal mechanisms and informing further scalability.

The evidence-based logic model will be used to structure the approach to quantitative and qualitative data collection, analysis methods and synthesis of the findings. Using a theory-based evaluation approach also provides the opportunity to deepen knowledge of how, and in what contexts, mediating variables individually and together interact with inputs, outputs and emerging outcomes, and in turn are impacted on, and changed by the trial.

The methods for gathering the qualitative data are outlined below, and will be mapped onto the logic model to ensure the research questions, methods and logic models are aligned throughout the evaluation process (see Table 1).

Research methods

9 school case study visits (3 in year 1 and 6 in year 2) will take place in schools receiving the NEBT intervention, in order to undertake the following data collection:

- Interview with the deliverer
- Observation of NEBT sessions
- Interview with the instructor
- Interview with the classroom teacher
- Focus group with approximately 8 Y5 pupils
- Interview (where possible) with member of SLT

The visits intend to understand if the programme was perceived to be effective from the deliverer's, instructors' and school recipients' perspectives. Visits would also attempt to identify some early programme impacts. Qualitative data collection will start with the deliverer interviews, before moving on to schools.

Piloting

Data collection tools for each of these methods will be developed and piloted wherever possible. For the case study visits, we will identify one NEBT intervention school suggested by ROE in advance of field visits, to check the coverage, appropriateness of questions and approaches, and duration (estimated timings are given below).

Cognitive testing of the terms/wording used, particularly for the pupil focus group, will be conducted in advance with similarly aged children. Consideration will be given to language and cultural sensitivity, flow of questions, time likely to be taken, understanding etc. Any changes or amendments will inform subsequent data collection.

Interview with the deliverer

Informed by the training and/or school visits to observe the NEBT delivery, online project manager interviews will be conducted in the autumn term of 2022, and summer terms of 2023 and 2024 to explore the:

- design and development of NEBT
- the recruitment of schools – was this as intended (deprivation criteria and compliance)
- the recruitment and training of instructors – was this as intended (compliance)
- support for and monitoring of instructor delivery (compliance and fidelity)
- session delivery - as intended (compliance and fidelity)
- factors influencing the delivery to date

- any changes in recruitment, instructor training and delivery from year 1 to year 2 (compliance and fidelity)
- delivery and attendance levels – minimum threshold for compliance/fidelity
- adaptions /considerations made to NEBT so sessions are inclusive for all pupils in terms of SEND, ethnic, language and cultural sensitivity (including the use of Canadian terms and lesson/educational norms and practices)
- any learning for next phase of recruitment and implementation – and extent of intended or unintended changes from year 1 to year 2
- perceptions of any outcomes and the effectiveness of the programme
- any unintended consequences / outcomes
- Scalability

School case study visits will be arranged during the spring and/or summer terms of each trial year to capture data after the NEBT sessions have become well established in the schools and at a point when teachers, instructors and pupils can reflect on the sessions to date and any outcomes so far.

Observation of NEBT sessions

Where possible in each year, two schools will be visited to observe a pre-family session, two visited during a family session, and two when a post-family visit session is underway. An observation schedule will be designed to capture key elements pertaining to session delivery, relative to the session plan (to assess the quality, compliance and fidelity), and pupil engagement and responses to the session. Cultural and racial appropriateness and sensitivities will also be noted. Observations will be followed up with interviews and focus groups, as outlined below.

Interview with NEBT instructor

This interview of up to 45 minutes would ask about the NEBT instructor's:

- other roles and experience of working in this or other schools and delivering similar projects or interventions to pupils
- usual practice – other social and emotional activities and support offered by the school and extent to which NEBT class experienced these during the intervention
- experience of recruitment by ROE/school
- NEBT training attendance and experience – was this as intended (quality, compliance and fidelity)
- ongoing ROE support throughout the delivery – was this as intended, has this been sufficient and helpful (quality, compliance and fidelity)
- recruiting and working with the family

- timetabling (timing and frequency) and preparation of sessions – as planned or adaptions made (compliance and fidelity)
- working with, and support from the class teacher and school
- delivery of the NEBT sessions to date – number/timing of sessions and extent of coverage of all aspects as intended (compliance and fidelity)
- instructor's log/notes from each session re coverage, issues arising and pupil attendance (compliance and fidelity) if available
- barriers and enablers of delivery - aspects that went well / not so well (quality and implementation, compliance and fidelity)
- pupils' engagement in each topic – appropriate pacing and coverage sessions pre family visit, during family visit and post family visit
- adaptations made for sessions to be inclusive of all pupils, in terms of SEND, ethnicity, language and cultural differences and sensitivities
- observations and perceptions about the specific affect, differences and impact of the family visit – compared to pre and post visit learning
- observations and perceptions of outcomes to date for different pupils and the wider class
- any improvements and learning about delivery and implementation

Interviews with the classroom teacher

During each case study visit, the Y5 class teacher will be interviewed for around 30 minutes to firstly understand the wider contextual issues that may have a moderating influence on the NEBT programme delivery and outcomes. This includes any relevant characteristics of the class in terms of their wellbeing and any socio-emotional or behavioural issues before the start of the NEBT programme.

In terms of understand usual practice, related PSHE curricular coverage and any other wellbeing or mental health initiatives, approaches or support provided in the school or class will also be explored to understand how this may affect the fit and impact of NEBT.

In relation to NEBT, we will focus on the class teacher's experience and perceptions of:

- usual practice
- the NEBT programme delivery by the instructor (quality, delivery, compliance and fidelity)
- timetabling of NEBT sessions (fidelity and compliance)
- coverage of topics and activities to date (fidelity and compliance)

- aspects of delivery have been effective/worked well or less well in their context (quality, delivery)
- pupils' attendance, engagement in and responses to sessions - how this varies for different pupils/sessions (quality, fidelity)
- observations and perceptions of the difference made by the family visit (quality, delivery, fidelity)
- cultural sensitivity and appropriateness of the content and delivery approaches
- any impacts to date in terms of pupil pro-social behaviour, emotional regulation, learning and how this varies across the class
- any unintended outcomes
- the data collection processes, teacher and pupil completion of the measures etc.

Focus group with approximately 8 Y5 pupils

As part of each case study visit, a Y5 pupil focus group will be conducted with about 8 pupils from the intervention class receiving the NEBT lessons. This would take up to 45 minutes and explore their perceptions, responses and experiences of:

- attendance and engagement with the NEBT programme, what they enjoyed or not (quality, delivery, fidelity)
- the topics covered in each of the elements/sessions – recollection and learning, e.g. about their emotions (quality, delivery, compliance, fidelity)
- their experience of having the family visit with the baby and parent – what this taught them and how it made a difference to their learning in the pre and post family visit sessions and overall
- any changes in how they feel and behave towards others at home and at school or noticed in others in their class
- changes in their relationships and friendships as a result of the NEBT lessons
- the completion of the pre and post measures/surveys – how easy/difficult they found it
- any other lessons, assemblies or support they have received over the academic year related to wellbeing, behaviour (e.g. PSHE, nurture groups etc) and how these have helped them deal with their emotions, friendships etc (usual practice)

Cards will be printed and used as prompts during the sessions. For example, with key words as reminders of some of the session topics. Interactive methods would be used to encourage all pupils to participate in the focus group discussion – e.g. for those who feel less confident in expressing their experience, asking them for individual words to describe their experience.

Interview with member of SLT

Where possible, a 20-minute interview will be arranged with the headteacher or other senior leader to understand:

- senior leader's reasons and motivations in taking part in the trial
- fit of the trial with other priorities, activities, curricula and approaches across the school e.g. PSHE/emotional wellbeing support (usual practice)
- ease of delivery for the school – additional resources, time/support needed for delivery (quality)
- barriers and enablers - what worked well (quality, delivery)
- observations and perceptions of the particular difference made by the family visit
- awareness of any impacts or outcomes to date – for different pupils
- any unintended consequences / outcomes

Racial and cultural sensitivity

The ROE intervention was originally intended to focus on the more disadvantaged areas of Doncaster, Birmingham, Northamptonshire, Nottingham and London (with geographical areas with over 21% Pupil Premium), but due to recruitment challenges, the areas have been extended more generally to Yorkshire, the Midlands, Merseyside and London (for cohort 1) and wider to include Wales in cohort 2. Although these are all areas of racial and cultural diversity, it is possible that the schools recruited (e.g. via networks, social media etc as well as targeting schools in these regions) may come from more advantaged areas than originally intended. As part of the analysis, we will collate school characteristic data (e.g. %FSM, %EAL, %SEN, KS2 attainment) to profile recruited schools and compare them with all schools in England.

At all stages of the evaluation, we will ensure that our approaches, including analysis, are inclusive and cognisant of all aspects of diversity. Once schools are identified from the MOUs, publicly available data will be used to collate a full profile of the schools' characteristics. This will enable the evaluation team to assess the representativeness and diversity of the schools in terms of their socio-economical, ethnic characteristics.

This information will be used to inform the case study interview questions (as outlined above), including asking school staff about the accessibility of the measures for Y5 pupils with different language and communication needs.

We recognise the limited collection of sensitive personal data at the pupil level to allow for a more granular quantitative analysis of pupil outcomes related to these characteristics. We

will therefore use the observations and case study qualitative data collection as opportunities to understand and critically consider issues of inclusivity. For example, observations of the instructor training and NEBT delivery in the classrooms will note the cultural appropriateness of the materials, language and approaches. More widely, this will include assessment of:

- the appropriate adaptation of generic ROE delivery for an English classroom context, given the Canadian genesis, development, and tone of the programme
- appropriate recognition and sensitivity of the content and delivery for a post-Covid classroom and school context - e.g., increased concerns about pupils' socio-emotional, wellbeing and behavioural issues in the ongoing Covid recovery period
- the language, materials, and delivery approaches from a trauma informed perspective, noting any sensitive and appropriate acknowledgement of the impact the pandemic, poverty, SEND, mental health and diverse family contexts may have on pupils, their learning, attachments and behaviours.
- explicit and implicit assumptions about family structure and relationships, e.g., language pertaining to traditional, white nuclear family structures, heteronormativity or ablism - or whether the language, delivery and resources indicate awareness and inclusion of other cultures, ethnicities and diverse family contexts and dynamics.

Interviews with the deliverers, instructors and teachers will also include questions around the racial and cultural appropriateness of the delivery model and thought given to inclusivity more generally (e.g. SEND, emotional needs and responses of pupils related to their family backgrounds, attachment styles, trauma etc). In FGs with pupils, evaluators will also explore the appropriateness of the delivery for their needs.

Bernadette Stiell who led the IPE qualitative strand for cohort 1 is from an ethnic minority background and has many years of professional and personal experience of equality, diversity and inclusion issues, having conducted a number of research projects with this as a main focus. She will be involved in the design and conduct of the qualitative research instruments and will ensure that the fieldwork team are briefed to explore issues across all aspects and stages of the evaluation. All members of the fieldwork team are also highly experienced in designing, conducting, analysing, and reporting IPE case study data, as well as working with primary school staff and pupils with a range of different needs. Pre-fieldwork briefings and analysis meetings will ensure the team have a shared and robust understanding of the data collection and analysis processes to ensure rigour and minimise any potential biases.

Qualitative analysis methods

All observations and interviews will be digitally recorded and transcribed for later thematic analysis. Analysis will specifically explore aspects of implementation, compliance, quality of delivery, fidelity, adaptation, responsiveness and usual practice. This will be analysed using a framework developed from the research questions above which map on to the coded logic model and theory of change. Codes and subcodes will be used to identify the inputs, activities, outputs and any perceptual evidence of outcomes, mediators and causal mechanisms or processes (i.e. identifying *how* the intervention works). Moderators and other contextual factors that modify the form or strength of the characteristics will also be taken into account, and as such also make up part of the coded logic model diagram (as shown in appendix 1).

Analysis in year 1 and year 2 will involve:

- Creating a case-data based approach to collate and code all qualitative data sources (observations, notes, interviews and focus groups), for thematic analysis using NVivo.
- An NVivo coding frame (based on Appendix 1) will be constructed to capture the key components of the:
    - logic model attributes – codes developed will relate to school and pupil characteristics, inputs, activities, outcomes, perceptions of outcomes, mediating factors (barriers and enablers), and any evidence relating to possible causal mechanisms or processes.
    - Trial - codes specifically related to aspects of the recruitment, training, implementation, quality of delivery, compliance, fidelity, adaptation, responsiveness, effectiveness and usual practice
    - Research questions mapped to the logic model - including changes from year 1 to year 2 perceptions and scalability
- Pre-analysis meetings will enable the qualitative team to co-construct the coding framework, have a shared and agreed understanding and opportunities to refine the codes and subcodes before the coding process and thematic analysis commences.
- Regular analysis meetings will check inter-rater reliability to ensure accuracy and consistency of interpretations, coding and thematic analysis across the cases
- Key themes and subthemes will be examined and compared as they emerge across the cases

Coding will therefore be primarily deductive, with some new top-level codes added inductively as they emerge from the data.

An interim report will be drafted after year 1 analysis has been completed, synthesising the findings from the qualitative analysis with the trial impact analysis by October 2023.

At the end of year 2, both sets of qualitative and quantitative findings will be combined and synthesised by December 2024.

When the quantitative and qualitative evidence from the two years is combined and compared, it will be used to test and update the logic model, identifying possible causal mechanisms and informing further scalability

Cost evaluation

The IPE will also collect cost details in each of the two years. The evaluation of the cost of the NEBT programme will follow YEF guidance and take a bottom-up approach. Cost details will include the cost of training and any costs incurred during the classroom delivery of the NEBT programme. Given that this evaluation will draw on two instances of the NEBT programme along with two impact evaluations and IPE, cost details will similarly be collected for both instances and the final cost evaluation will combine these as an average. Estimated total cost along with cost per pupil participant will be provided.

The organisations and practitioners involved in delivery are both Roots of Empathy and intervention schools. To calculate costs, we will be collecting data on costs associated with the delivery of the programme for both RoE and intervention schools, split between phase (start-up, pre-requisite and recurring) and based on actual costs, and time taken and for whom. We will collect information on activities that are needed for delivery, but that do not have a cost attached, for example, volunteering. Below is an indication of the core activities/costs we envisage being aspects of NEBT delivery, and for whom.

- Training and recruitment of instructors, including instructor applications, meetings, preparation time for, and attending interviews etc (RoE & Schools)
- Cost of RoE package, without discount (School)
- Training, ongoing CPD and mentoring (School)
- Time for mother and baby recruitment (School & RoE)
- Preparation of lesson plans, delivery of lessons (Schools)
- Completion of RoE programme forms and feedback (Schools)
- Instructor top-up training (School & RoE)
- Printing and/or photocopying costs
- Any unexpected costs

Collection of cost data will be carried out both as part of a school survey, and during the IPE data collection and interviews/conversations with Roots of Empathy. We will survey all intervention schools to collect the above listed data, and to determine the role of the member of staff (e.g. TA, SENCO) trained as a RoE instructor, this will then be used to create an average cost for school-based staff wages. This will ensure that we have a clear understanding of how much variation in cost would occur as a result of different members of staff trained to be a RoE instructors. Interviews with school staff and RoE will focus on establishing whether the components we have selected are the main components with costs attached, how much RoE's delivery elements cost etc.

Ethics and registration

SHU has undergone a full review and approval processes through the university ethics committees (Ref: ER19810112. This involved writing a detailed application that was reviewed by Sheffield Hallam University independent ethics reviewers. This trial has been registered at the ISRCTN (International Standard Randomised Controlled Trial Number) and the registration number is ISRCTN98490275.

Data protection

SHU and ROE will strictly comply with current data protection legislation, including the GDPR. Under GDPR Article 6, Paragraph 1e, the legal basis for this project is it being a 'public task', as the research is being conducted to evaluate the impact of an approach to building social emotional skills and empathy that has potential benefits for pupils participating in the trial and beyond. However, because of future matching of trial data to the, NPD and PNC we have decided to collect parental opt-out consent for participation. In addition, parents/carers are free to withdraw their child from data collection and analysis at any time until the data is archived at the end of the project. Information on how to withdraw will be provided for schools, parents and carers. If a parent/carer decides to withdraw, this would mean that no data on their child would be included in the evaluation and the child would not be required to take the measures (surveys) but can still participate in the NEBT sessions. If you would like further information, please contact the delivery team at ROE using the contact details at the end of this document.

SHU are the Data Controllers for the data collected as part of the Nurturing Empathy before Transition project evaluation, up until the evaluation has finished. After the evaluation is finished (in 2024), the pupil data collected will be sent to the Department for Education (DfE) (at which point SHU cease to be responsible for the data), where it will be pseudonymised and transferred to the secure archive, which is being held by the Office for National Statistics in their Secure Research Service (SRS). Once the data has been transferred to the SRS, the Youth Endowment Fund become responsible for the data.

No pupils will be individually identifiable in the data archived and archived data will be kept indefinitely. Further information on YEF's data archive can be found below.

A data sharing agreement and fair processing notice will detail the personal data to be shared, and a fair processing notice will be sent to all participating schools as per GDPR requirements.

Stakeholders and interests

Evaluation team: Sheffield Hallam University

Dr Sarah Reaney-Wood: PI and impact evaluation lead. Sarah went on maternity leave in January 2024 but will contribute to the write up of the report in Autumn 2024.

Whilst Sarah is on maternity leave, the impact evaluation will be led by Sean Demack who will be supported by Giota Blouchou.

For cohort 1, the implementation and process evaluation (IPE) was led jointly by Bernie Stiell and Claire Wolstenholme with support from Lucy Clague. Following Bernie Steill's departure from SHU, the IPE for cohort 2 is now led jointly by Josephine Booth and Eleanor Byrne.

The extent of change to the evaluation team has been notable; a total of eight SHU academics have been involved during the five years of the project with two of these people (Sean Demack and Sarah Reaney-Wood) having consistent involvement. Some of this can be accounted for by the timescale of the project and the impact of Covid 19. SHU were appointed as evaluators for the first round of YEF projects 2019 but the evaluation was postponed because of the Covid 19 pandemic so that recruitment took place in the 2021/22 academic year. The post-pandemic difficulties in recruitment led to the decision to adopt a split-cohort design with cohort 1 running in 2022/23 and cohort 2 running in 2023/24. This meant that the initial plans for a two-year project and one-year evaluation in 2020/21 (recruitment in 2019/20) expanded to a five year project with two one-year evaluation cohorts; cohort 1 in 2022/23 (recruitment in 2021/22) and cohort 2 in 2023/24 (recruitment in 2022/23).

Developers & delivery team: Roots of Empathy

Katie Cohen will lead the NEBT programme at ROE.

Emily Standfield is the manager of research for ROE.

Risks

RISK DESCRIPTION AND IMPACT	MITIGATIONS	REVISED RISK CATEGORY	REVISED IMPACT LEVEL	REVISED PROBABILITY LEVEL
Difficult to recruit the number of schools needed for the trial. This will have an impact on the viability of the trial and the evaluation.	A long lead in time has been arranged to allow for RoE to have a period of 'soft-recruitment', and a period of formal recruitment. RoE are recruiting in 4? geographical areas, which increases the number of potential schools.	Low	High	Low
Schools in more disadvantaged areas may face more challenges in participating in the trial.	Incentives for participating are being given to schools. Regular contact with intervention and control schools throughout the process (emails, phone calls) to inform, remind, encourage, support, allay fears wherever possible so potential issues can be identified and addressed early. Flexibility and accommodation of challenges within the timeframe and methodology of the trial.	Medium	Medium	Medium

Schools do not complete baseline data to acceptable threshold in time for randomisation deadline. Impact would be that we have unbalanced samples and missing data.	Inclusion threshold of 60% completion of data. Ensure schools have enough time for data completion, an awareness of deadline for completion, and that non-completion will result in them not being part of the trail. Schools will receive their randomisation allocation soon after data completion. Participation is incentivised. Professional services staff to chase schools with missing data.	Medium	Medium	Medium
Instructor training only taking place in two locations may lead to restricted attendance from instructors and possible school dropout.	RoE are supporting schools around instructor recruitment, recognising the capacity issues schools have in releasing/covering staff. SHU providing information to schools about process to allay fears. ROE to pay for transport and accommodation for instructors to attend training.	Medium	Medium	Medium
Unsuccessful instructor interviews may lead to a reduction in the quality and integrity of RoE's programme	RoE interviewing and supporting schools through their application and screening process.	High	High	Low

Attrition at the school and/or pupil level due to longitudinal data matching	For all YEF trials, the data collected will be archived to be matched to the PNC to allow longitudinal tracking of outcomes. Schools/parents may feel uncomfortable with this level of data matching, and therefore either decide to withdraw from the trial (school level) or not consent for their child to take part (individual level). The reason for longitudinal data matching is clearly defined in the MoU and the information sheets, in addition parents/carers are made aware that the data is anonymised. Staff at SHU are available during the recruitment stages to talk through any concerns that might be raised about data matching.	Medium	High	Low
Staff absence/departure (e.g. due to long term illness)	Researchers in the SIOE with experience on YEF/EEF trails would be able to cover staff absences or departures.. The centre has a low staff turnover and the same team see projects through from inception to completion in almost every instance.	Low	Low	Low
Young people reluctant or unable to fully engage in the data collection activities	The evaluation team will employ a toolkit of methods for data collection that would maximise the likelihood of CYP engaging. The pupils of focus in this trial are Y5. The questionnaires that have been chosen as the	Low/Medium	Medium	Low/Medium

	primary and secondary outcome measures have been carefully selected with this age group in mind, they are validated with this age-group In addition, we are only collecting these outcome measures at two time-points, keeping burden to a minimum. For the IPE data collection with pupils, we will utilise learning from other projects (such as creative data collection methods) to improve engagement.			
The impact of Covid on pupils' emotional wellbeing poses additional challenges for the intervention compared with previous evaluations.	Both Covid and the cost-of-living crisis will be taken into account and the team will include questions around this in the case study fieldwork. School staff (leaders, teachers, and instructors) and deliverer (ROE PM) interviews and pupils focus groups will explore the appropriateness, suitability and effectiveness of the NEBT in the current context of Covid impacts and challenges.	high	medium	medium
Further Covid 19-related disruption	Team will closely monitor and follow government guidelines around safe working. Staff are able to work remotely, offering flexible remote fieldwork options where possible. Range of virtual methodologies available to use with participants.	Medium	Medium	Medium

Programme delivery issues	As the NEBT programme is centred around the parent-baby relationship and requires one visit per month from parent and baby into schools, there is the potential that issues such as: illness, baby getting upset etc may have an impact on programme delivery. Case studies will explore the extent to which this was an issue and how it was dealt with.	Medium	Low	Medium
End point data collection issues.	Control schools are paid £200 on completion of the baseline data, and a further £200 for endpoint data collection, thus incentivising completion. Regular communication with control schools will ensure they have information, reminders, and support in completing this at each stage.			

Timeline

Dates (WHEN)	Activity (WHAT)	Staff responsible/ leading (WHO)
April-July 2022	School recruitment for cohort 1	RoE
September-October 2022	Baseline data collection for cohort 1 schools (class lists, pupil survey, teacher survey)	SHU
September-October 2022	Randomisation for cohort 1	SHU
October 2022	Cohort 1 intervention delivery starts	RoE
October 2022	Control schools receive part of their incentive payment	SHU
October 2022 – May 2023	IPE school visits for cohort 1	SHU
May-June 2023	Intervention delivery ends	RoE
June 2023	Cohort 1 endpoint data collection (pupil survey & teacher survey)	SHU
Sept 2022 - July 2023	School recruitment for cohort 2	RoE
September 2022	Evaluation protocol published.	SHU
September-October 2023	Baseline data collection for cohort 2 schools (class lists, pupil survey, teacher survey)	SHU
September-October 2023	Randomisation for cohort 2	SHU

October 2023	Cohort 2 intervention delivery starts	RoE
Feb - May 2023	IPE school visits for cohort 2	SHU
March 2024	Evaluation Statistical Analysis Plan published	SHU
April 2024	Control schools receive incentive payment 1	SHU
May-June 2024	Intervention delivery ends in Y2 schools	RoE
June/July 2023	Endpoint data collection in Y2 schools	SHU
July-Sept 2024	Analysis	SHU
August 2024 – November 2024	Write up	SHU
Sept 2024	Control schools receive incentive payment 2	SHU
December 2024- March 2025	Peer review of final report and revisions	YEF & SHU
March-April 2025	Report shared with ROE and updates / revisions	YEF, SHU & ROE
July 2025	Final report published	YEF

Publication

A key feature of YEF's publication policy is that the first report published about the impact of the intervention will be the evaluation report. (6.-YEF-publication-policy.pdf

(cloudinary.com)) As such, no other publications will be published until the impact report has been published by YEF.

References

Browne G., Gafni A., Roberts J., Byrne C., Majumdar B. (2004). Effective/efficient mental health programs for school-age children: a synthesis of reviews. *Soc Sci Med* 2004;58:1367–84. https://doi.org/10.1016/S0277-9536(03)00332-0

Connolly, P., Miller, S., Kee, F., Sloan, S., Gildea, A., McIntosh, E., Boyer, N., & Bland, M. (2018). A cluster randomised controlled trial and evaluation and cost-effectiveness analysis of the Roots of Empathy schools-based programme for improving social and emotional well-being outcomes among 8- to 9-year-olds in Northern Ireland. *NIHR Journals Library.*

Deighton, J., Tymms, P., Vostanis, P., Belsky, J., Fonagy, P., Brown, A., Martin, A., Patalay, P. & Wolpert, W. (2013). The Development of a School-Based Measure of Child Mental Health. *Journal of Psychoeducational Assessment*, 31:247.

Durlak J. A., Weissberg R. P., Dymnicki A. B., Taylor R. D., Schellinger K. B. (2011). The impact of enhancing students' social and emotional learning: a meta-analysis of school-based universal interventions. *Child Dev* 2011;82:405–32. https://doi.org/10.1111/j.1467–8624.2010.01564.

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40(11), 1337-1345.

Hamoda, H. M., Chiumento, A., Alonge, O., Hamdani, S. U., Saeed, K., Wissow, L., & Rahman, A. (2021). Addressing the consequences of the COVID-19 lockdown for children's mental health: Investing in school mental health programs. *Psychiatric services,* 72(6), 729-731.

Haylock, S., Boshari, T., Alexander, E.C., Kumar, A., Manikam, L., & Pinder, R. (2020). Risk factors associated with knife-crime in United Kingdom among young people aged 10–24 years: a systematic review. BMC Public Health, 20, 1451.. https://doi.org/10.1186/s12889-020-09498-4

Jolliffe, D., & Farrington, D. P. (2006). Development and validation of the Basic Empathy Scale. *Journal of adolescence*, 29(4), 589-611.

Latsch, D., Stauffer, M., & Bollinger, M. (2017). Evaluation of the Roots of Empathy program in Switzerland, years 2015 to 2017. Full Report. Bern: Bern University of Applied Science.

Lee, J. (2020). Mental health effects of school closures during COVID-19. *The Lancet Child & Adolescent Health*, 4(6), 421.

Office for Health Improvement and Disparities. (2022,). Covid-19 mental health and wellbeing surveillance: Report. GOV.UK. Retrieved May 6, 2022, from https://www.gov.uk/government/publications/covid-19-mental-health-and-wellbeing-surveillance-report

Panayiotou, M., Humphrey, N., & Wigelsworth, M. (2019). An empirical basis for linking social and emotional learning to academic performance. *Contemporary Educational Psychology,* 56, 193-204.

Santos, R. G., Chartier, M. J., Whalen, J. C., Chateau, D., & Boyd, L. (2011). Effectiveness of school- based violence prevention for children and youth: Cluster randomized field trial of the Roots of Empathy program with replication and three-year follow-up. *Healthcare Quarterly,* 14, 80-91. http://www.longwoods.com/articles/images/HQ_vol14_ChildHealth_Issue2_Santos.pdf

Tomé, G., Almeida, A., Ramiro, L., & Gaspar, T. (2021). Intervention in Schools promoting mental health and well-being: a systematic review. *Global Journal of Community Psychology Practice,* 12(1).

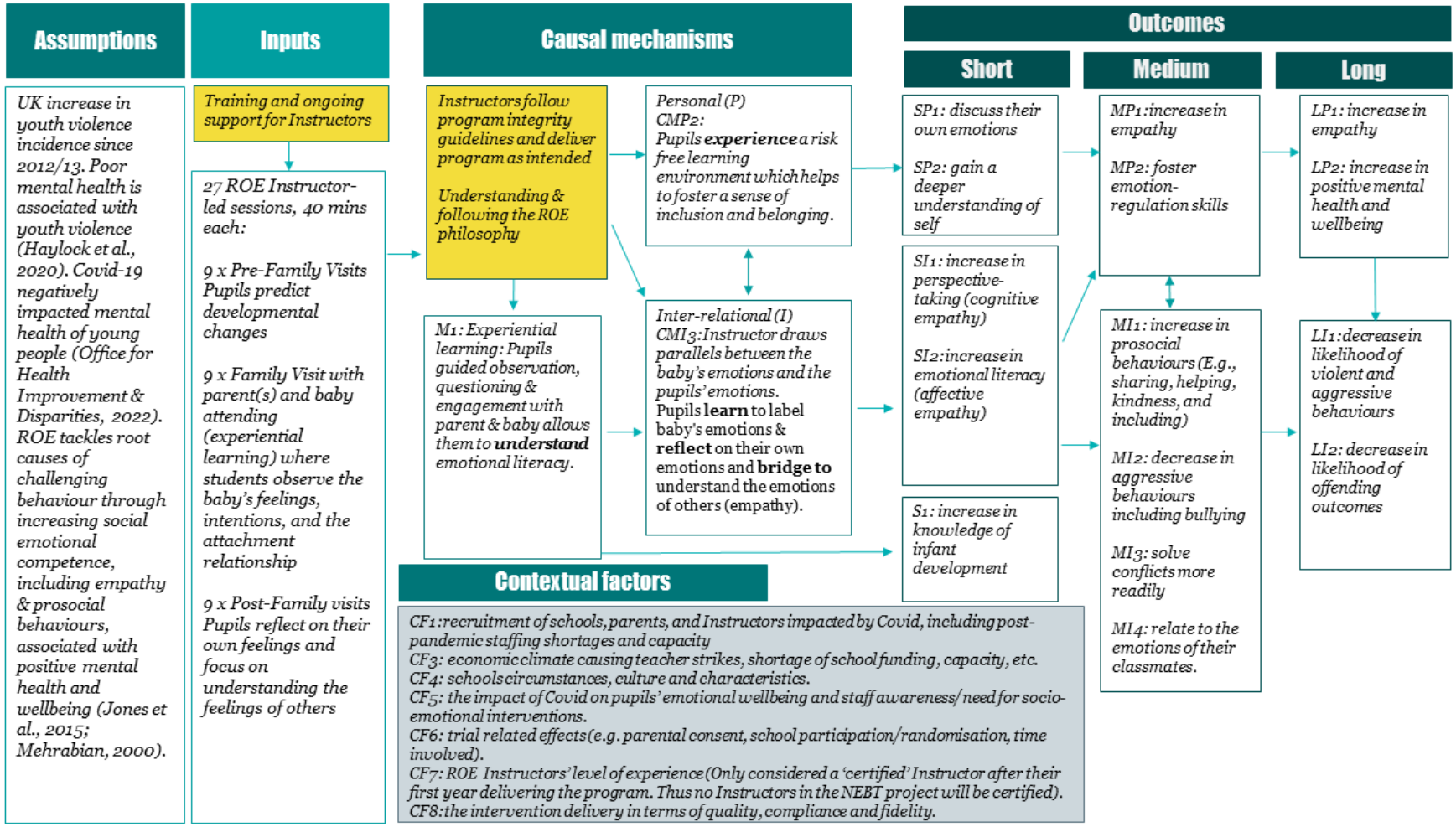
Wrigley, J., Makara K., Elliot D., (2015). Evaluation of Roots of Empathy in Scotland 2014-2015, Final Report for Action for Children (unpublished).

Appendix 1: Updated Impact and IPE activity and how this links to RQs and LM

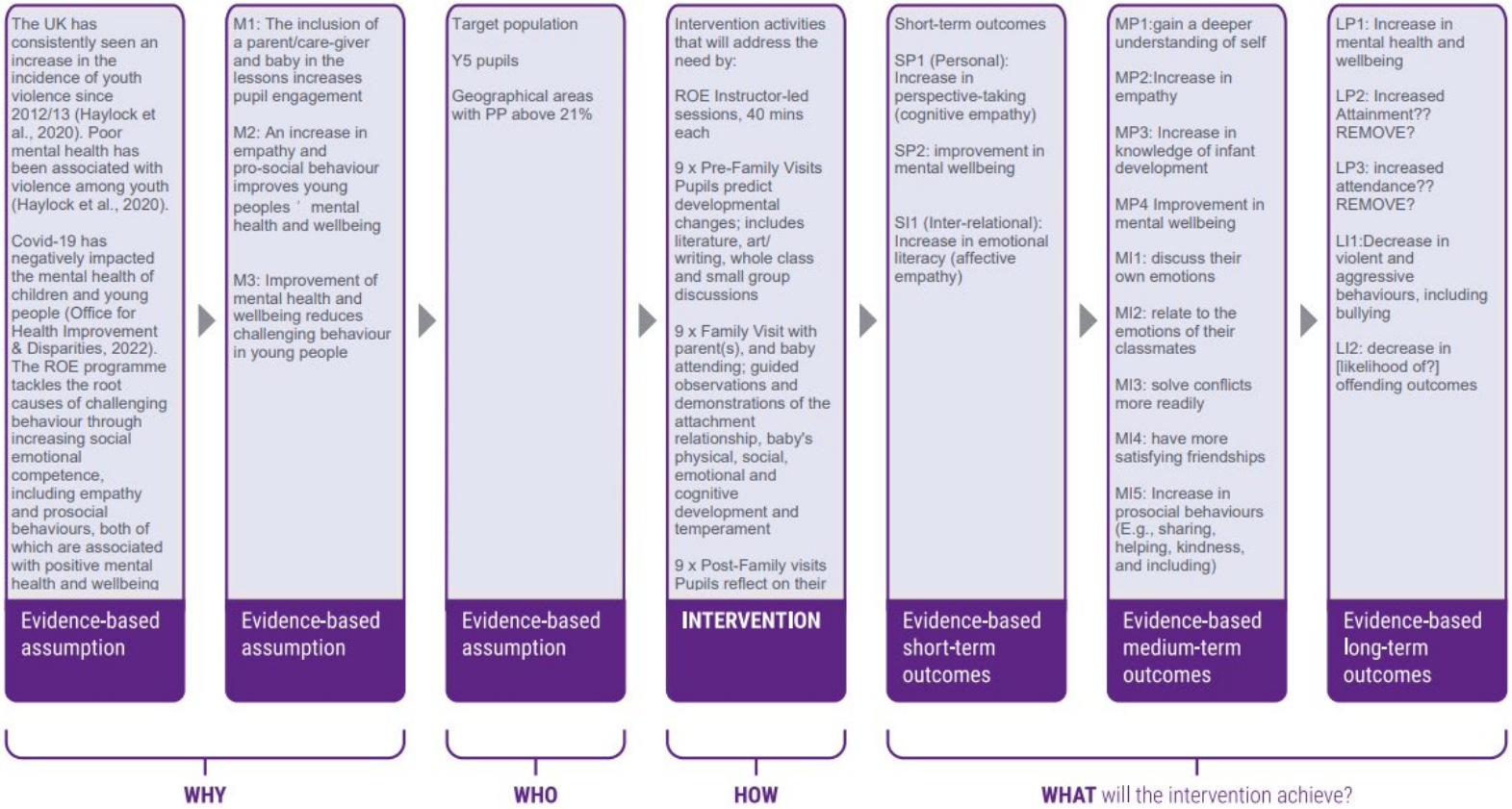
Impact or IPE	Data collection method	Outcome/Measure	Research Question	Logic model code
Impact evaluation	Pupil Survey at baseline & outcome in both cohorts.	Two Me and My feelings measures: Behaviour Diffs & Emotional Diffs (pre and post)	RQ1-2	SP1, SP2, MP2, MI2, MI4, LP2
	Pupil Survey at baseline & outcome in both cohorts.	Two BES measure Affective Empathy & Cognitive Empathy (pre and post)	RQ3-4	SI1, SI2 MP2, MI1, MI3, MI4 LP1
	Teacher Survey at baseline & outcome in both cohorts.	Six Teacher SDQ measures Total Diffs, Prosocial, Emotional Probs, Conduct Probs, Peer Probs, Hyperactivity (pre and post)	RQ5-10	MI1, MI2, MI3, MI4, LI1, LI2
Implementation and Process evaluation (IPE)	Case studies (overall)	3 case studies in Y1 (summer 2023) to reflect smaller cohort size 6 case studies in Y2 [3 in Wales and 3 in England] (summer 2024)	RQa-d	CM1, CMP2, CMI3 SP1, SP2, SI1, SI2, S1 MP1, MP2, MI1, MI2, MI3, MI4 CF1-8
	Observations	Instructor Training observations: Y1: 1 visit Oct 2022 Y2: 1 mid-year visit in Feb 2024 (no data observation) Classroom observations of ROE sessions: Y1: 2 in summer 2023 Y2: 6 in summer 2023	RQc RQa,c	(yellow inputs & yellow causal mechanisms) CM1, CMP2, CMI3 SP1, SP2, SI1, SI2, S1, CF8

	NEBT Instructor interviews	Y1: 3 interviews Y2: 6 interviews	RQa, RQb, RQc	(yellow inputs & yellow causal mechanisms) CM1, CMP2, CMI3 SP1, SP2, SI1, SI2, S1 MP1, MP2, MI1, MI2, MI3, MI4 CF1-8
	Class teacher interviews	Y1: 3 interviews Y2: 6 interviews	RQa, RQb, RQc	CM1, CMP2, CMI3 SP1, SP2, SI1, SI2, S1 CF3-6, CF8 MP1, MP2, MI1, MI2, MI3, MI4
	SLT interviews	Y1: 3 interviews Y2: 6 interviews	RQa, RQb, RQc	CF1-8 Possibly: SP1, SP2, SI1, SI2, S1 MP1, MP2, MI1, MI2, MI3, MI4
	Pupil focus groups	Y1: 2 focus groups Y2: 6 focus groups Approx 8 pupils in each group	RQa, RQb, RQc	CM1, CMP2, CMI3 SP1, SP2, SI1, SI2, S1 MP1, MP2, MI1, MI2, MI3, MI4
	Deliverer interviews	Y1: 1 PM/deliverer interview Y2: 1 PM/Deliverer Interview	RQa, RQb, RQc, RQd	Yellow inputs; CF1-8 (and their perspectives on CMs and outcomes)
	IPE analysis insights		RQa, RQb, RQc, RQd	All LM

Appendix 2: Finalised NEBT logic model



Appendix 3: EIF template logic model



YOUTH ENDOWMENT FUND

Impetus

Home Office

youthendowmentfund.org.uk

hello@youthendowmentfund.org.uk

@YouthEndowFund

The Youth Endowment Fund Charitable Trust

Registered Charity Number: 1185413