

Trial Evaluation Protocol

Level 4 Group Triple P : Positive Parenting Program®

Evaluator (institution): RAND Europe

Principal investigator(s): Elena Rosa Brown



Template last updated: March 2018

PROJECT TITLE	Level 4 Group Triple P
DEVELOPER (INSTITUTION)	<i>Parenting and Family Support Centre, University of Queensland</i>
EVALUATOR (INSTITUTION)	RAND Europe
PRINCIPAL INVESTIGATOR(S)	Elena Rosa Brown
PROTOCOL AUTHOR(S)	Dr. Sashka Dimova, Amelia Harshfield, Dr. Andreas Culora, Dr. Alex Sutherland, Elena Rosa Brown, Natalie Picken
TRIAL DESIGN	Two-arm cluster randomised controlled trial with random allocation at the Early Years setting level
PUPIL AGE RANGE AND KEY STAGE	3 and 4, Early Years (Nursery)
NUMBER OF SCHOOLS	150
NUMBER OF PUPILS	1,800
PRIMARY OUTCOME	Expressive language
SECONDARY OUTCOME	Child behaviour

Protocol version history

VERSION	DATE	REASON FOR REVISION
1.0		

Table of contents

Protocol version history.....	1
Table of contents	2
Intervention.....	3
Study rationale and background	6
Impact Evaluation.....	7
Research questions.....	7
Design	8
Randomisation.....	8
Participants.....	9
Sample size calculations	11
Outcome measures.....	14
Analysis plan	15
SENSITIVITY ANALYSIS	17
Implementation and process evaluation	17
Cost evaluation	22
Data protection.....	22
Personnel	23
Risks	23
Timeline	24
Appendix A: Description of what is covered in each session with parents	26
APPENDIX B: DESCRIPTION OF BASELINE DATA COLLECTION PROCESS.....	27
References	28

Trial Evaluation Protocol

Level 4 Group Triple P : Positive Parenting Program®

Evaluator (institution): RAND Europe

Principal investigator(s): Elena Rosa Brown



Template last updated: March 2018

Intervention

Name: Level 4 Group Triple P – Positive Parenting Program®

Parenting programmes are among the most promising strategies for improving child wellbeing through reducing child behavioural problems, inconsistent and harsh parenting, child maltreatment, and promoting positive parenting practices (World Health Organization, 2016). Converging evidence suggests that parenting programmes can be equally effective for the most disadvantaged families (Bröning et al., 2017; Gardner et al., 2017).

Positive parenting has been identified as a key protective factor for healthy development (Denham et al., 2000; Gardner et al. 1999; Hutchings et al. 2007). Conversely, behavioural problems in childhood are particularly associated with hostile, critical, punitive and coercive parenting (Rutter et al. ,1998). The Triple P Positive Parenting Program® is a system of parenting interventions focused on developing positive parenting skills and techniques, including parental self-efficacy and self-management, through five core principles of positive parenting. These principles are: safe and engaging environment, positive learning environment, assertive discipline, realistic expectations and parental self-care (Sanders, 2008). It was developed by Professor Matt Sanders and colleagues at The University of Queensland in Australia (Sanders, 1999) and is currently disseminated globally by Triple P International, based in Brisbane, Australia (Triple P Positive Parenting Program). Triple P encompasses multiple interventions targeting different ages and populations, and has five different intensity levels, all with the common purpose of enhancing the knowledge, skills and confidence of parents (Sanders, 2008). Level 1 is a communication strategy designed to reach all parents, Level 2 is a one-to-three-session seminar to parents of children aged up to 12 years old and teens with mild behaviour problems, and Levels 3-5 are weekly sessions of varying time-periods, which typically target parents of children with behavioural difficulties (Sanders et al., 2014; Nowak & Heinrichs, 2008). This evaluation focuses on the delivery of Level 4 Group Triple P, for parents of 3 to 4 year old children with language difficulties, family vulnerability, and/or severe behavioural difficulties.

Triple P draws from several theoretical principles. These include social learning models of parent-child interactions (Patterson, 1982), child and family behaviour therapy and applied behaviour analysis (Risley, 1976), developmental research on social and intellectual competence in early parent-child relationships (Hart & Risley, 1995), risk and protective factors and developmental psychopathology (Rutter, 1985; Patterson, 1982), cognitive social learning theory (Bandura, 1977 & 1995), and public health and community psychology (Sanders, 1999).

Of particular relevance to this project is that Triple P draws heavily upon the work of Hart and Risley in the parental influence on child language and communication (Hart & Risley, 1995). The programme strongly emphasises incidental teaching and promoting natural use of language at home. Children are more likely to develop their language ability in a safe, positive, low-conflict environment where they feel comfortable initiating conversation (Payne et al., 1994). Its core principles were selected from the developmental literature to address specific modifiable risk and protective factors known to predict positive child developmental and mental health outcomes. During Triple P programmes, parents learn strategies to apply these principles to their family interactions. This project offers an opportunity to identify the specific effects of the intervention on language outcomes because, at present, there is no direct evidence of language outcomes relating to Triple P.

In this evaluation, Group Triple P will be delivered to parents of 3 and 4 year old children in approximately 50% of 150 early years settings in the north of England. Group Triple P will be implemented by multi-disciplinary practitioners working in these early years settings who have at least an NVQ level 3 qualification or a higher qualification in health or education, early childhood education or social services. Triple P practitioners are nominated practitioners from these settings who have attended training and have become accredited to deliver the programme to parents. Triple P training addresses the importance of delivery fidelity whilst also exploring permissible flexibility. Triple P practitioners are therefore expected to be highly skilled in the flexible delivery of Triple P to suit different family needs, while maintaining programme fidelity for successful outcomes. It is recommended that practitioners will be Early Years (EY) practitioners or, ideally, senior EY practitioners (due to the high level of expertise required).

Early years settings will be recruited by the Triple P team but will be allocated to either the intervention or control group by RAND Europe.

The implementation of Triple P will involve the following activities (see **Figure 1. Triple P Logic Model** below):

- **Infrastructure**

- **Practitioners' training:** Two practitioners¹ from each setting will receive training on how to deliver the programme between November 2019 and January 2020. Training will be delivered by Triple P (TP) trainers.² Training consists of the following elements:
 - **Triple P training:** Typically around twenty practitioners will participate in a three-day course delivered by one TP trainer.
 - **Pre-accreditation workshop:** Triple P training will be followed by a one-day pre-accreditation workshop designed to support preparation for accreditation.
 - **Accreditation** The half-day accreditation workshop is facilitated by a different TP trainer, meaning that each practitioner will have worked with two TP trainers. Practitioners will prepare for their accreditation by completing a multiple-choice quiz ahead of their accreditation date and rehearsing their role-plays with the other practitioner from the same setting. Accreditation is through The University of Queensland and requires a demonstration of skills.
 - **Post-accreditation:** In the last week of January 2020, practitioners will attend a one-day clinical workshop delivered by a TP trainer which further prepares the practitioner for Triple P delivery by offering skills refinement and maximising confidence in programme content and process.
- **Triple P Sessions:** Trained practitioners will commence programme delivery in the first week of February 2020. This will include eight weekly sessions with maximum of 12 parents. Detailed information for each session is given Appendix A The first four sessions will be delivered face-to-face as group sessions.
- These four group sessions will be followed by three one-to-one practical and personalised telephone consultations.
- Finally, there will be one face-to-face group session, which will complete the programme and parents' contact with the Triple P practitioners. The main aim of this session is to review progress and plan for the future.
- Group sessions will each last approximately two hours, while the telephone consultations will take approximately fifteen to thirty minutes.
- The face-to-face group meetings will be held at multiple settings, including community centres and nursery schools and other early years settings.

¹ Training is offered to two practitioners per setting in order to mitigate risk of attrition. However, it is not mandatory criteria for settings to have two practitioners at the training.

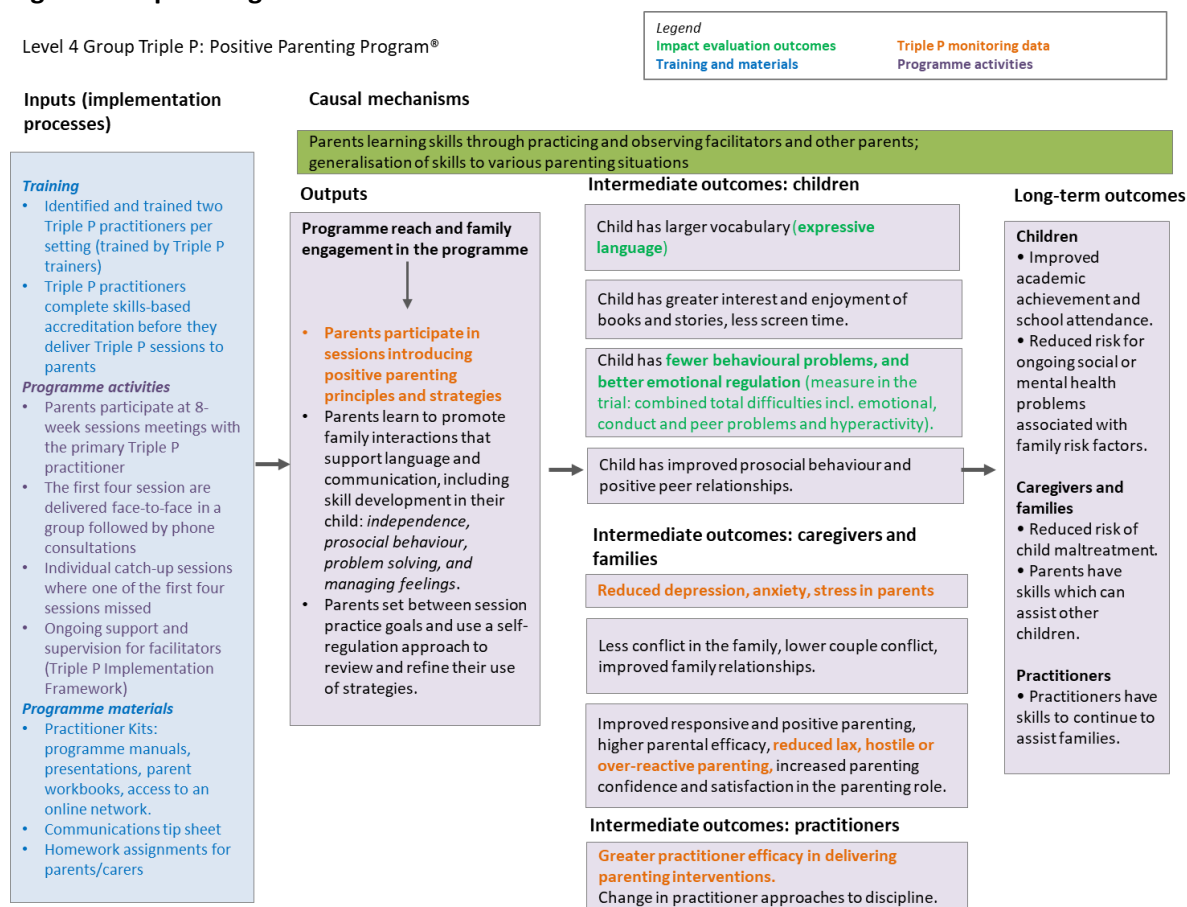
² Triple P trainers are trained facilitators, experienced and accredited contractors or members of TP staff, typically holding psychology qualifications (usually MSc or CPsychol).

- The group sessions will usually be delivered while children are attending care in the same setting (during daytime on weekdays). [Practitioners will be able to decide when sessions will be delivered and may schedule them at the evenings or weekends. However, they will be advised that attendance will be highest if childcare is provided].
- Practitioners will be instructed to schedule individual catch-up sessions if the parent misses one of the first four sessions.
- **Support** will be provided for practitioners in early years settings in the intervention group. Support includes telephone or video call supervision by Triple P trainers to practitioners with the goal to support programme adoption, delivery with fidelity and sustainability to the programme. This implementation support will also focus on how to tailor the programme and processes to suit the local context. For this project, an emphasis on the aspects of Triple P related to language and communication development will be included in both training and ongoing implementation support. To match the difficulties presented by parents, practitioners will give communication examples for each strategy discussed during the sessions throughout programme delivery. They can introduce one or more Communication Tip Sheets, which will support parent homework following Session 2 of Triple P and will reinforce the strategies discussed.

Triple P programmes are manualised, meaning that practitioners follow manuals with detailed information on the programme content and principles when implementing the programme. Practitioners will receive Facilitator Kits for Group Triple P and a Communication Tip Sheet. These kits contain the programme manuals, copies of the parent workbooks, Presentation CD and DVD (Group Triple P), and Communication Tip Sheet. Practitioners will also receive access to the Triple P Provider Network where they can download assessment measures and access FAQs. Practitioners will need laptops and data projectors for programme delivery. Parents will receive programme parent workbooks and a communication tip sheet.

Early years settings in the intervention group will not receive any monetary payment. However, each will receive training and accreditation in Level 4 Group Triple P for two members of staff, and be provided with the resources required to deliver the programme to parents in the trial without any of the usual associated costs (usually approximately £3,500 per site for two practitioners per site). Settings assigned to the control group will not take part in Level 4 Group Triple P. However, they will receive a payment of £750 either on 30 September 2020 or once post-programme testing is completed (depending on which is sooner) that will be dependent on the setting taking part in the post-test data collection. Settings in the control group must not deliver any parenting programmes up until this point of payment.

Figure 1. Triple P Logic Model



Study rationale and background

Recent reviews of early intervention, such as those by Allen (2011a) and the National Academy for Parenting Research (Asmussen & Weizel, 2010), provide increasingly clear and objective advice on a range of effective family-based programmes aimed at improving parenting.

A broad body of well-designed studies has tested the effectiveness of these family-based programmes and demonstrated a positive impact on parenting skills and children's behaviour (Gould et al., 2006). Programmes have also been noted to have a positive effect on parental mental health (Lindsay et al., 2011), improve children's school attainment (Scott et al., 2010) and reduce the number of children placed on Child Protection Registers and in local authority care (Prinz et al., 2009).

Level 4 Group Triple P has been extensively evaluated through RCTs (Leung et al. 2003; Crisante et al. 2003; Zubrick et al., 2005; Chung et al. 2015; Kim et al. 2018; Smith et al., 2018). Positive results on the effectiveness of the Triple P programme have been reported regarding parenting skills, child problem behaviour, and parental well-being (Nowak & Heinrichs, 2008; Early Intervention Foundation, 2018; Doyle et al 2018).

Similarly, a meta-analysis on 101 empirical studies of the evaluation of Triple P (at all levels) concluded that the programme can be effective in reducing children's behavioural problems, reducing parents' dysfunctional parenting practices (Sander et al., 2014). However, the methodological quality of the studies included in this meta-analysis was not discussed.

Still, a few studies, including an independent evaluation in the United Kingdom (UK), reported null effects on child and parent behaviours (e.g. an RCT of the Level 4 Group Triple P with 146 families in Birmingham by Little et al., 2012). Furthermore, a meta-analysis review based on 32 evaluations on Triple P programmes of various levels³, found that the selected evaluations in the same were small including fewer than 35 participants per study arm (Coyne & Kwakkenbos, 2013; Hoath & Sanders, 2002 Wilson et al., 2012) and lacking follow-up, due to wait-list designs where active treatment was offered to control settings following the active post-intervention data collection (Wilson et al., 2012).

The largest population study and RCT is the U.S. Triple P System Population Trial, which involved 18 counties being randomly assigned to either the Triple P Positive Parenting Programme or the services-as-usual control services. It is estimated that approximately 8 to 13 thousand families participated in the population study. The study found those counties who were assigned to one of the Triple P programmes demonstrated improvements in population indicators related to child maltreatment⁴ (Prinz et. al, 2009). Additionally, Level 4 Group Triple P has been researched at large scale as part of previous Government rollouts of evidence-based parenting programmes and as part of subsequent on-going regular service delivery in England, and has demonstrated significant improvements in child behavioural issues, as well as significant improvements in parental styles and parental well-being, as well as maintenance of those effects at 12-month follow up (Lindsay & Strand 2013; Gray et al, 2018). The largest randomised controlled trial study in Europe was run in 56 Swiss elementary schools, involving around 1600 children entering first year at elementary school. The intention to treat analysis in this study did not find that participating in the intervention decreased children's externalising problem behaviour (Malti et al., 2011).

Most of the evidence around Triple P is from evaluations that were conducted outside the UK, all of which measure behavioural or parenting outcomes, rather than the impact on children's language learning outcomes. The effect of Triple P on children's language is therefore unknown; hence the need for this study. There is substantial value in understanding whether a programme primarily intended to improve family relationships and behaviour can have an impact on children's language acquisition, as measured through expressive language. To our knowledge, this will be one of the largest independent evaluations of Triple P in the UK in terms of the number of families included in the study, and the first study of its kind to test effects on language.

This efficacy trial will help determine whether Level 4 Group Triple P for 3 and 4 year olds leads to observable improved outcomes in children's expressive language. The current evaluation is funded by the Education Endowment Foundation (EEF), Department for Education (DfE) and SHINE Trust. The programme is delivered by the Triple P UK team and the trial will be carried out by RAND Europe, who are independent evaluators appointed by the EEF.

Impact Evaluation

Research questions

The core research questions this project seeks to answer are:

1. To what extent did Triple P lead to changes in children's expressive language outcomes?

³ The evidence in the meta-analysis is not limited to Level 4 Group Triple P. It includes evidence from various Triple P programmes.

⁴ Population indicators include: substantiated child maltreatment, child out-of-home placement, and child maltreatment injuries.

2. To what extent did Triple P lead to changes in children’s behavioural outcomes?

The impact evaluation is designed to investigate the following research hypotheses:

H1. Children in the intervention group, Triple P, will have higher levels of expressive language compared to children in the control (business as usual) condition.

H2. Children in the intervention group, Triple P, will have fewer behaviour problems compared to children in the control (business as usual) condition.

Design

Trial type and number of arms		Two-group, parallel, stratified, cluster-randomised controlled trial (cRCT) at the setting level
Unit of randomisation		Early years settings
Stratification variables (if applicable)		Cluster 1: Newcastle 1; Cluster 2: Wakefield; Cluster 3: Manchester 1; Cluster 4: Manchester 2; Cluster 5: Newcastle 2; Cluster 6: Preston; Cluster 7: Liverpool. ⁵
Primary outcome	variable	Expressive language
	measure (instrument, scale)	Composite expressive language score made from the Clinical Evaluation of Language Fundamentals (CELF) Level 1 ⁶ subtests: 1. Sentence structure 2. Word structure 3. Expressive vocabulary
Secondary outcome(s)	variable(s)	Child behaviour
	measure(s) (instrument, scale)	The full Strengths and Difficulties Questionnaire Teacher (SDQ-T). ⁷

The Triple P evaluation will be a two-group parallel, stratified, cluster-randomised control trial, with early years settings being the unit of randomisation, and child outcomes as the unit of analysis. To ensure comparability of these settings across the intervention arm and the control arm (‘exchangeability’, ⁸ see Oakes 2013), we will randomise within the different geographical areas; Doing so serves to balance study arms on geographical location, and therefore any regional differences. Using geographical area will allow us to have similar number of intervention and control settings in each region.

Randomisation

Randomisation will be conducted using Stata by a member of the evaluation team in October 2019.

⁵ Note the clusters are defined based on the location for training.

⁶ Wiig, E. H., Secord, W., & Semel, E. (2000). Clinical Evaluation of Language Fundamentals (CELF)—Preschool UK Edition.

⁷ Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40(11), 1337-1345.

⁸ That is – if one were to swap the allocation of settings to intervention and control groups, the results from the trial should be the same.

In preparation for randomisation, we will examine the distribution of schools by region and type of nursery (school-based vs. private, voluntary and independent nurseries (PVI)). The location for training will be the main stratifying variable, while stratification by type of nursery will be incorporated to ensure that PVI settings were not allocated to treatment or control unevenly. To deal with unequal treatment fractions we will use the command `randtreat` and the option `misfits(global)` in Stata (Carril, 2017).

Settings are expected to be recruited from the following regions: Greater Manchester, Liverpool City region; North and South Tyneside; Newcastle upon Tyne; County Durham; Northumberland; North, South, West and East Yorkshire; Cheshire East and Cheshire West and Cheshire and Blackpool, Cumbria, Lancashire.

The recruited settings will be organised in seven different clusters depending on the location for the practitioner training. The following training clusters have been identified: Cluster 1: Newcastle 1; Cluster 2: Wakefield; Cluster 3 Manchester 1; Cluster 4: Manchester 2; Cluster 5: Newcastle 2; Cluster 6: Preston; Cluster 7: Liverpool

Table 1. Description of regions within training cluster

Training Cluster	Regions
Cluster 1: Newcastle 1	North Tyneside, Northumberland, County Durham, Newcastle, Cumbria
Cluster 2: Wakefield	North, South, West and East Yorkshire
Cluster 3 Manchester 1	Greater Manchester
Cluster 4: Manchester 2	Greater Manchester, Cheshire East and Cheshire West and Cheshire
Cluster 5: Newcastle 2	Northumberland, Cumbria, County Durham, South Tyneside
Cluster 6: Preston	Lancashire, Blackpool
Cluster 7: Liverpool	Liverpool, Lancashire, Cheshire, Greater Manchester.

Randomisation will occur in October 2019, with allocation revealed to settings once each has completed teacher SDQ and parent information baseline data has been collected. Settings will be required to provide the evaluation team with baseline data to be considered part of the trial. If they fail to do so, they will not be randomised and included in the trial. The trial allocation will be recorded and communicated to the Triple P team and the EEF in an Excel file that will be password protected to prevent editing.

Baseline equivalence will be examined based on the initial randomisation. A well-conducted randomisation will, we expect, yield groups that are equivalent at baseline (Glennister & Takavarasha, 2013). Because settings will be randomly allocated to the control and intervention conditions, any imbalance at baseline will have occurred by chance. To assess imbalance at baseline, we will compare groups at setting and child levels, by means of cross-tabulations and histograms that assess the distribution of each characteristic within the control and intervention groups (EEF, 2018).

SELECTION OF SUB-SAMPLE OF PARTICIPATING FAMILIES FOR OUTCOME TESTING

A maximum of twelve families per setting will be selected to participate in the Triple P study. If a setting has more than ten participating families, RAND Europe will randomly select ten families in each setting to take part in the evaluation. These will be selected from the lists of eligible families that will be provided by the settings at the beginning of the study.⁹ If a control group setting recruits twelve families, then the two “extra” families will not end up being part of the study at all. If an intervention group setting recruits twelve families, then the two “extra” families will be part of the intervention but will not complete outcome testing.

Participants

SETTINGS

⁹ Families will be sorted by a random variable in stata and the first 10 families after ranking the random number will be included in the pool for randomisation

The Triple P team will aim to recruit early years settings that are school-based nurseries in the first instance. However if they find this challenging, they will also consider recruiting non-school-based nurseries into the trial: including private, voluntary and independent nurseries (PVI) or standalone state nurseries. However, as there is a risk that non-school-based nurseries will be harder to follow up and would have higher attrition, it was agreed that the number of such schools will be limited in the trial to 20% of the total cohort, ideally spread across different areas. The following eligibility criteria will be used for recruitment for all settings:

- The setting is located in the areas for recruitment determined by the EEF and the Delivery team¹⁰
- The setting has not had Triple P or Incredible Years¹¹ delivered to 3 and 4 year olds since 1 January 2018.
- The setting is not actively involved in any other EEF Home Learning Environment Trial.
- The settings must be willing to:
 - provide background information to the delivery team (as specified in the Memorandum of Understanding);
 - release two practitioners to take part in the training and deliver Triple P to parents;
 - support the administration and collection of tests two times within the project's timeline (the SDQ-T questionnaire to be completed by the teachers and administered at the beginning and end of the study, while child's language assessment will be administered at the end of the study);
 - be randomly assigned to intervention or 'business as usual' at the setting level;
 - engage with the delivery team and implement the intervention;
 - facilitate data collection by the evaluation team.

Preference will be given to settings located in areas of high disadvantage.¹² Participating early years settings will sign a Memorandum of Understanding (MoU) which will outline the roles and responsibilities of all stakeholders involved. Settings and participating families will be asked to consent to sharing the specified data with the delivery and evaluation team. Settings will be asked to notify the delivery and/or research team immediately if a practitioner or family withdraws from data collection.

CHILDREN

Before randomisation, teachers will nominate up to 12 children with language delay or other reported concerns around their behavioural, emotional and/or social development who are willing to take part in an eight-week Level 4 Group Triple P parenting programme in Spring 2020. Children will be eligible if:

- They are 3 or 4 years old; AND
- The teacher is concerned about a delay in a child's language / communication development. OR

¹⁰ Eligible areas are:

Greater Manchester, Liverpool City region; North and South Tyneside; Newcastle upon Tyne; County Durham; Northumberland; North, South, West and East Yorkshire; Cheshire East and Cheshire West and Cheshire and Blackpool, Cumbria, Lancashire.

¹¹ This is another evidence based parenting programme evaluated in numerous studies in the UK (for more information look at <http://www.incredibleyears.com/>)

¹² 'Easy' ways of doing this are where settings are located in Lower-layer Super Output Areas (LSOA) with higher than average levels of deprivation compared to the rest of the region as measured by the Index of Multiple Deprivation (IMD).

- The teacher or parent is concerned about a child's behavioural, emotional or social development¹³; OR
- The practitioners decide that there is an indicator of increased family vulnerability, such as: family social disadvantage, financial stress, housing insecurity, adverse life events.

In many early years settings, there is a home visit by setting staff before the start of the academic year, meaning that practitioners will have a better idea about the family's situation. Many early years staff will also have "at-the-door" conversations when children are picked up/dropped off where parents may also discuss any concerns about the child.

PARENTS/CARERS

Parents (or carers) of nominated children will be eligible if:

- They understand and speak English; AND
- They have not attended *either* Triple P or Incredible Years parenting programmes since January 2018; AND
- They agree that they and their child can participate in the research and that they will attend the eight-week programme in the Spring term 2020, if allocated to the intervention group.

The primary caregiver (e.g. the parent spending the majority of time with the child) is advised to attend the sessions. In usual practice, *both* parents are encouraged to attend and if one is not able to attend, the other is encouraged to tell the other parent.

Parents with severe psychological illness cannot be excluded prior to initial contact with practitioner, as the practitioner may be unaware of their psychological state. However, it is suggested that their inclusion is avoided if this information later comes to light as these parents may have difficulties engaging in the programme. Both the delivery and evaluation team have agreed, however, that it is not necessary to specify illness as a separate criterion as those with a severe mental or physical health problem may not be available and willing to attend.

Sample size calculations

		ICC=0.11	ICC=0.13	Early Year Pupil Premium
Minimum Detectable Effect Size (MDES)		0.194	0.203	0.271
Pre-test/ post-test correlations	level 1 (pupil)	0.50	0.50	0.50
	level 2 (class)	N/A	N/A	N/A
	level 3 (setting)	0.00	0.00	0.00
Intraclass correlations (ICCs)	level 2 (class)	N/A	N/A	N/A
	level 3 (setting)	0.11	0.13	0.13
Alpha		0.05	0.05	0.05

¹³ The rationale for mentioning parents in regard to behaviour is that child behaviour is expected to sometimes vary a lot between home and settings, whereas language development will be more constant, and therefore less important to include both parent and teacher perspectives. The general approach is that teachers make the decisions, but can incorporate parental concerns.

Power		0.8	0.8	0.8
One-sided or two-sided?		Two	Two	Two
Average cluster size		10	10	3
Number of schools	Intervention	75	75	75
	Control	75	75	75
	Total	150	150	150
Number of pupils	Intervention	750	750	225
	Control	750	750	225
	Total	1500	1500	450

Our assumption is that this trial should be powered to detect an effect of $d=0.20$ or smaller based on the primary expressive language outcome measures described below. We believe it would be important to power to this level, even though this is an efficacy trial, because of the questions over intervention effects and the distal nature of the expressive language outcome (as the primary target of the programme are behavioural outcomes as opposed to other parenting programmes primarily targeting language development, e.g., Mol et al, 2009).

In previous early years research, setting-level attrition has been estimated to be as high as 15% (Robinson-Smith et al., 2018). On the basis of the above factors *we propose recruiting 150 settings*, which builds in a 'buffer' for attrition at person and setting levels and allows for some variability in the intra-cluster correlation (ICC). Power and minimum detectable effect size (MDES) calculations were performed using the PowerUp tools for main effects (Dong & Maynard, 2013) and moderators (Spybrook et al., 2016).

The amount of variation explained is assumed to be 0.25 (equivalent to correlation of 0.50) for level 1 (pupils) and 0.00 for level 2) (based on Charman et al., 2015). We present the MDES estimates for two different intra-cluster correlations (ICCs) (ICC= 0.11 and ICC=0.13) to illustrate the impact of these on the MDES. We assume an average cluster size of 10 pupils (3 and 4 year old children) in each early year setting. We also assume an alpha of 5% and an intended 80% power to detect effects. We use two-level clustered designs, assuming a continuous, normally distributed (Gaussian) outcome.

Using the parameters above and with equal allocation to intervention and control the MDES is 0.194 or 0.203. As such, even though considered an efficacy trial, the study should be powered to detect meaningful differences between groups assuming there are 150 settings.

Assuming that on average there are around three children eligible for Early Years Pupil Premium (EYPP) with an ICC of 0.13, the MDES is 0.271.

Update: Due to recruitment challenges we expect lower number of recruited settings and participants. The number of settings in the study it is expected to be between 95 and 100. Furthermore, the number of participants per setting is expected to range between four and twelve. Under more conservative assumption, with 96 settings and eight participants per setting, the MDES is estimated to range between 0.254 and 0.266 for ICC=0.11 and ICC=0.13, respectively. The risk is that the trial will be underpowered to find an effect smaller than 0.26. It has been agreed in consultation with Triple P UK and the EEF that the larger MDES will be reasonable as this is a targeted intervention.

		ICC=0.11	ICC=0.13	PP
Minimum Detectable Effect Size (MDES)		0.254	0.266	0.390
Pre-test/ post-test correlations	level 1 (pupil)	0.50	0.50	0.50
	level 2 (class)	N/A	N/A	N/A
	level 3 (setting)	0.00	0.00	0.00
Intracluster correlations (ICCs)	level 2 (class)	N/A	N/A	N/A
	level 3 (setting)	0.11	0.13	0.13
Alpha		0.05	0.05	0.05
Power		0.8	0.8	0.8
One-sided or two-sided?		Two	Two	Two
Average cluster size		8	8	2
Number of schools	Intervention	48	48	48
	Control	48	48	48
	Total	96	96	96
Number of pupils	Intervention	384	384	96
	Control	384	384	96
	Total	768	768	192

Outcome measures

PRIMARY OUTCOME

The primary outcome for this evaluation is the child's expressive language. The rationale for having a particular focus on expressive language is two-fold. Existing research shows that challenges for lower-attainment language learners are particularly significant in the area of expressive skills and it is more sensitive to changes than receptive language (Gibson et al., 2012; Rogde et al., 2016).

To measure expressive language, we will use the Clinical Evaluation of Language Fundamentals for Preschool, Second Edition (CELF-P2). The general expressive language ability will be based on the level 1 core language score of the CELF-P2. This is a measure of general language ability based on the following subtests: (i) Sentence Structure; (ii) Word Structure; (iii) Expressive Vocabulary¹⁴ (Wiig, E. et. al., 2006). The core score is derived by summing the scaled scores from each of the three subtests and converting the overall score to standard score¹⁵. However, we will also present results from each component by treatment and control groups to assess the relative importance of each component. All sub-test component measures are standardised in the UK, are age-appropriate and were selected by the evaluation and delivery team in collaboration with language experts advising the evaluation team.

The primary outcome measure will be administered at the end of the trial in the week commencing 22 June 2020. Post-intervention outcome data will be collected by Elklan¹⁶ testers working for the evaluation team, who will not have knowledge of the allocation for each setting. This approach allows for blinding to allocation: we can supply a list of settings to Elklan without revealing allocation. This is valuable given the risk of bias introduced by having assessors who are aware of treatment allocation (Higgins et al., 2016).

In order to ensure a minimum of 90% completion rates and limit the amount of missing data. Elklan shall conduct two rounds of testing sessions irrespective of allocation to condition arm: first, to test all pupils; second, to test pupils missed in the first round. Testing will be completed in three weeks, including one week of mop-up testing.

SETTINGS WITH MORE THAN 10 CHILDREN¹⁷

To reduce the testing burden and costs, it was agreed to administer the CELF-P2 to ten randomly selected children.

Intervention setting: If the setting is selected to deliver Triple P, two nominated practitioners at the setting will receive the Triple P manual, training, resources and support needed to deliver the programme. The programme will be delivered to all parents, but only ten randomly selected children will be tested with the CELF-P2.

Control setting: Control settings will continue implementing business as usual reading practices. If the setting has selected more than ten children at the beginning of the programme only ten randomly selected children will undertake post-trial outcome testing.

¹⁴ The Sentence Structure test is based on 22 items and 0 to 1 scale; the Word Structure consists of 24 items and 0 to 1 scale, while the Expressive Language score has 20 items and 0 to 2 scale, where 0 is given for incorrect and 1 or 2 for correct response.

¹⁵ An overall score can be converted to standard score by using the Core Language and Index Standard Scores in the appendix C of the CELF manual. For more information see: Wiig, E., W. Secord, and E. Semel. "Clinical Evaluation of Language Fundamentals—Preschool—2, UK Standardisation (CELF-P2 UK)." (2006).

¹⁶ Elklan have experience and expertise in testing young children to ensure age-appropriate and engaging outcome measurement. They have a large network of language specialists who are experienced in working with young children, minimising recruitment delay and overall travel time/costs. <https://www.elklan.co.uk/>

¹⁷ If the setting selects 10 or less children, all children will complete the CELF-P2

SECONDARY OUTCOME

The secondary outcome will be child behaviour, as this is a key outcome measures in other trials of Triple P (De Graaf et al., 2008; Sanders et al., 2003), and the working assumption is that both behaviour and expressive language should improve, as set out in the preceding paragraph and our hypotheses. We propose surveying all practitioners in order to gain data on child behaviour in early years settings. As agreed with Triple P UK, we will use the Strength and Difficulties Questionnaire Teacher (SDQ-T) to assess child behaviour focusing on the Total Difficulties score (Goodman 1997). The SDQ-T is a well-known, short and structured instrument measuring child behaviours and can be used by parents and educators. It has 25 items each rated on three point scale (not true, somewhat true and certainly true). The total difficulty score can range between 0 and 40 and is based on the answers of all scales (emotional problem scale; conduct problem scale; hyperactivity scale; peer problem scale) except the prosocial scale. The teacher version SDQ- T has been validated as reliable instrument for identifying behavioural problems in children aged 3-4 years, including children born in the United Kingdom (it was based on data from 16 659 families) (Croft et al., 2015; Ezpeleta et al, 2013).

Both baseline and post-trial SDQ-T will be collected for the children selected into the study before randomisation (based on the criteria above)¹⁸. Baseline testing will be completed before randomisation, while post-trial SDQ –T will be collected in June 2020. Collection of baseline data is mandatory for selected settings and families to be considered part of the trial. At baseline, the SDQ-T will be administered by practitioners within the chosen early years settings. Baseline data collection will be conducted by early years setting practitioners to reduce evaluation costs. As this will take place pre-randomisation, this data will be unbiased by knowledge of allocation. As the existing online platforms for SDQ scoring can be cumbersome and require a log-in, we have decided to set up the SDQ on more user-friendly survey platform.¹⁹ Additionally, if practitioners find it easier to read/respond to items on paper, we will provide them with paper versions. Practitioners can therefore fill in either a paper or online version of the SDQ questionnaire. Although the scoring is simple, it will be easy to make mistakes when scoring for many children at the same time, which is not ideal for baseline score precision. To minimise mistakes the online platform will automatically calculate the total difficulty score after the items are completed. For more information on the process please see **Error! Reference source not found.** in the **Appendix B**.

In the summer term 2020, both control and intervention settings will be asked to repeat the SDQ-T for those children that were originally selected. The staff will also receive online surveys.

OPTIONAL FOLLOW-UP

There is a pressing gap in the literature of examining the potential long-term impacts of early years interventions (Sim et al., 2018). This trial would offer a good opportunity to administer a follow-up testing six months or more after the end of the intervention, once pupils have transitioned into Reception or Year 1. This is currently being considered by the EEF and thus not discussed further in this protocol.

Analysis plan

The outcome analysis will be on an intention-to-treat (ITT) basis. Once randomised, early years settings and children will be analysed according to the allocation of the setting regardless of whether the early years setting complied with the intervention or not. The analysis will include all randomised schools/pupils in the groups to which they were randomly assigned, regardless of: the treatment actually received, withdrawal from the intervention post-randomisation, or deviations in programme implementation. This principle is key in ensuring an unbiased analysis of intervention effects. This approach compares outcome means for the treatment and comparison groups, and subjects are analysed according to their randomised group allocation. The ITT approach is inherently conservative as it captures the averaged effect of offering the intervention.

¹⁸ Note if there are more than 10 families selected to participate in the Triple P programme, the SDQ T will be collected for all children but only 10 randomly selected children will be selected to sit the CELF test

¹⁹ We will use the Smartsurvey's platform.

The primary outcome for the 3 and 4 year old children will be expressive language as measured by the core language score from the CELF-Preschool 2 UK assessment. The core language score will be standardised with mean of zero and standard deviation of one. To estimate the impact on the primary outcome we will use a two-level multilevel model to account for clustering of data. Multilevel approaches assume that the schools in the study are a random sample of all schools and that the multilevel modelling framework can flexibly handle complex variation within/between schools (Snijders, 2015; Snijders & Bosker, 1994).

The general equation for the multilevel model is given below as Eq.(1):

$$y_{ij} = \alpha + X_{ij}\beta + Z_jb_j + \delta TripleP_j + u_{ij} + u_j \quad i = 1..N, j = 1..M, \quad (1)$$

where y_{ij} denotes the pupil level outcome; i and j denote child and setting indexes, respectively; X_{ij} is the $1 \times k$ vector of individual characteristics that include SDQ-T as a pre-test; Z_j is a vector of the stratification variables mentioned above (region); $TripleP_j$ is a dummy variable denoting intervention/control group at the setting level; β and δ are the $k \times 1$ and 1×1 vectors of regression coefficients; u_{ij} is the child-level error term; and u_j is the setting-level error term. The coefficient δ will constitute the main result of the trial that will tell us the average effect of the intervention on pupil outcomes in treatment schools compared to those in control schools.

SECONDARY OUTCOME ANALYSIS

The secondary outcomes will be based on the SDQ T difficulty score. Child' behavioural problems will be assessed following the same specification to equation (1) listed under primary outcome analysis above, but we will substitute these measures as outcomes.

SUBGROUP ANALYSIS

The study will not be powered to detect significant differences between sub-groups. However, we will report mean outcomes by sub-categories (EYPP eligible children) as a basic descriptive step. We will also conduct an exploratory analysis of differential effects for EYPP eligible children as they are considered a key target group by the EEF²⁰. As an exploratory modelling approach, EYPP will be incorporated into the regression analysis as a binary variable [1] if a child is considered eligible for EYPP and as [0] if they are not. The EYPP indicator will then be interacted with treatment allocation to assess the conditional impact of Triple P on EYPP eligible pupils.

IMBALANCE AT BASELINE

We will take an active approach to address imbalance by stratifying the randomisation. Balance checks will be conducted at the setting and child level.

At the setting level, we will check the balance in the following variables by means of cross-tabulations and histograms that assess the distribution of each characteristic within the control and intervention groups aggregated from the data in the study sample:

- Average baseline SDQ T score
- Proportion of children eligible for EYPP in the setting.

At the individual (pupil) level, balance will be assessed for the following characteristics:

- EYPP status for pupils
- Gender
- Average age in months

²⁰ FSM pupils are key considered key target group by the EEF. However, we are collecting the EYPP eligibility data from the settings as children can claim FSM only when they are 4. EYPP data for the current year will be collected directly from the settings.

- SDQ T baseline score

EFFECT SIZE

We will use the effect sizes for cluster-randomised trials given in the EEF evaluator guidance – an example, adapted from Hedges is given below:

$$(1) ES = \frac{(\bar{Y}_T - \bar{Y}_C)_{adjusted}}{\sqrt{\sigma_S^2 + \sigma_{error}^2}}$$

Where $(\bar{Y}_T - \bar{Y}_C)_{adjusted}$ is the mean difference between intervention groups adjusted for baseline characteristics and $\sqrt{\sigma_S^2 + \sigma_{error}^2}$ is an estimate of the population standard deviation (variance). The effect size therefore represents the proportion of the population standard deviation attributable to the intervention (Hutchison et al., 2010). The exact effect size used will depend on whether there are equal or unequal sample sizes in trial arms.

The same approach will be used for primary and secondary outcomes.

MISSING DATA

With respect to missing data, attrition across trial arms will be explored as a basic step to assess bias (Higgins et al., 2011). Systematic differences between dropouts and those retained will be modelled using missingness at follow-up as an outcome, predicted by baseline covariates including treatment. For item non-response, the extent of missingness in part determines the analytical approach. For less than 5% missingness overall a complete-case analysis should suffice, regardless of missingness mechanism (EEF, 2018). Our default would be to check results using approaches that account for missingness, relying on the weaker MAR assumption and building missing at random (MAR) conditioning variables from our initial work predicting missingness. For example, if there was systematic missingness in predictor variables, we would explore options for using FIML and/or multiple imputation (EEF, 2018; for a discussion of FIML vs MI see Allison, 2012).

SENSITIVITY ANALYSIS

As a sensitivity analysis we will undertake multilevel modelling for clustered designs for each component of the core language score. Child's language skills will be assessed by substituting these measures as outcomes into equation 1. The pretest score will be the covariate, and the posttest score the outcome measure. All analyses will be performed in Stata, versions 15.1 onwards (Heß, 2017).

Implementation and process evaluation

During kick-off meetings and the Intervention Delivery and Evaluation Analysis (IDEA) workshop, all parties worked to develop a detailed theory of change (TOC). The main goals of the meeting were to finalise the EEF's Template for Intervention Description and Replication (TIDieR) framework, and to discuss the logic model and how the Implementation and Process Evaluation (IPE) data was to be collected at each stage.

The purpose of the process evaluation is to address the following questions:

RQ 1: Was the intervention implemented with fidelity in the settings allocated to Triple P condition?

RQ 2: What are the barriers, facilitators and conditions needed to make Triple P succeed?

RQ 3: What are drivers of impact? What are the necessary conditions for success of the programme in terms of achieving impact?

RQ 4: What does "business as usual" look like in control settings?

Table 2 presents an overview of the data collection activities that will be part of the evaluation alongside the specific aims of these activities. Further details on the data collection methods are discussed in further detail below.

Table 2: Data collection activities and aims

Phase	Activities	Aims
Pre-randomisation	<ul style="list-style-type: none"> • IDEA workshop (RAND Europe) • Review of Triple P training materials (RAND Europe) • Collect information on reason for selection of children into the trial (RAND Europe) 	<ul style="list-style-type: none"> • Development and refinement of TOC (RQ 1, 2, 3, 4) • Documenting the quality of the intervention (RQ 3, 4) • Documenting the quality of the intervention (RQ 2)
Delivery period (ongoing 2019/2020)	<ul style="list-style-type: none"> • Collect practitioner questionnaires at 4 time points: pre- and post-training, pre- and post-accreditation (Triple P team) • Collect parent attendance logs (RAND Europe with practitioners) • Collect parent measures administered by practitioners in treatment arm (Triple P team) • Collect parent views of behaviour via text messages to parents in control arm (RAND Europe) 	<ul style="list-style-type: none"> • Documenting quality of and practitioner responsiveness to intervention (RQ 1, 2) • Documenting dosage of intervention/compliance (RQ 2, 3) • Documenting child behaviour in the home in the treatment arm (RQ 1, 2) • Documenting child behaviour in the home in the control arm (RQ 1, 2, 4)
Post-delivery period	<ul style="list-style-type: none"> • Practitioner survey in treatment arm (RAND Europe) • Practitioner survey in control arm (RAND Europe) • Phone interviews with parents (RAND Europe) • Online survey of practitioners to collect costs (RAND Europe) 	<ul style="list-style-type: none"> • Exploring experience of settings in delivery and impact (RQ 1, 2, 3) • Documenting counterfactual and any interventions that may impact on child behaviour or language (RQ 4) • Exploring experience of parents and parent responsiveness to intervention (RQ 1, 2, 3) • Reporting on costs associated with delivering Triple P (cost section of report)

1. PRE INTERVENTION PHASE (REVIEW OF MANUAL, REASON FOR SELECTING CHILDREN/FAMILIES INTO THE TRIAL)

RAND will review the training materials to build better understanding of the programme and its key components, which will help inform the later IPE tasks, such as the survey/interview questions.

Selecting the appropriate children is a key part of the Theory of Change in terms of leading to better outcomes. Before randomisation practitioners will be required to document the reason(s) why they have selected the child into the trial.

2. DATA PRODUCED BY TRIPLE P

As part of usual practice Triple P collects a rich amount of data from practitioners and parents. These include:

1. Practitioner questionnaires on confidence at 4 time points: pre- and post-training and at pre and post-accreditation
2. Aggregated parent measures administered by practitioners in treatment arm: Parenting scale, DASS 21 (adult Depression, Anxiety and Stress Scale 21) and parents SDQ²¹ (child behaviour)

Triple P UK will share data with RAND Europe in aggregated form as this is a cost-effective approach that will allow us to use data from all the settings. The practitioners confidence rating will be aggregated at cluster level. The delivery team will aggregate the parent measures at setting or cluster level.

Parent SDQ data from the treatment arm will be compared to results from texts to parents in control arm (see ATTENDANCE LOGS)

ATTENDANCE AT PROGRAMME SESSIONS is a key part of the TOC in terms of leading to better outcomes. To minimise the number of missed sessions they will be delivered whilst children are attending the setting (during daytime on weekends). Parents' attendance for each session will be documented by practitioners delivering Triple P. The attendance logs will be designed by the Evaluation team and they will be easy to use. Attendance at sessions will be used in defining the compliance measure.

TEXTS TO PARENTS IN CONTROL ARM). To minimise error and missing data entry, parent data will be scored and inputted centrally, after practitioners supply the data to a central UK point.

Practitioner questionnaires will be administered at 4 different time points: pre- and post-training, pre- and post-accreditation. Quantitative data from the questionnaires (e.g. practitioner confidence, parent attendance) will be summarised with results used to provide a description of how the intervention worked in practice. This will give an insight at how effective training is at equipping practitioners to deliver the programme.

3. ATTENDANCE LOGS

Attendance at programme sessions is a key part of the TOC in terms of leading to better outcomes. To minimise the number of missed sessions they will be delivered whilst children are attending the setting (during daytime on weekends). Parents' attendance for each session will be documented by practitioners delivering Triple P. The attendance logs will be designed by the Evaluation team and they will be easy to use. Attendance at sessions will be used in defining the compliance measure.

²¹ Parent SDQ is filled by the parent or carer of the child, while the teacher SDQ (SDQ-T) is filled by the teacher who knows the child best

4. TEXTS TO PARENTS IN CONTROL ARM

To compliment the parental measures being administered by Triple P practitioners to parents in the treatment arm, RAND Europe will send light-touch text surveys²² to all parents in the control arm. These texts will seek to understand child problem behaviour in the home and will be taken from or similar to an item in the parent SDQ (e.g. “has your child done Y in the past two weeks”). The other reason for collecting this information is to ensure that research contact is equalised between treatment arms in the trial, as asymmetrical 'attention' may otherwise confound later analyses.

These responses will be collected at two time points aligned to when the parent SDQ is being collected in the treatment group (i.e. pre-intervention and 8 weeks later). Responses will be scored in line with SDQ guidelines and presented descriptively in the final report.

Results will then be used to generate an understanding of child behaviour in the home across both treatment and control arms and to make note of any differences. This will further be used to examine the logic model (discussed above).

5. PRACTITIONER SURVEYS

We propose to conduct online surveys with all practitioners in both arms of the trial at the end of the academic year. These will be designed to collect perspectives of how the intervention has been implemented (in the treatment arm) and to understand what activities were conducted as business-as-usual (in the control arm)²³. We will include questions in the end-line survey in both arms to collect information on how settings i) work with parents, ii) work with parents to improve behaviour and language in the home, and iii) details of how these activities may or may not have differed in the evaluation year compared to previous years. Whilst it is important to receive feedback from treatment settings, we would prioritise understanding the activities of control settings as there will be good implementation data collected by Triple P.

Descriptive findings from the survey will be aggregated and summarised, with results used to provide a description of how the intervention worked in practice. Additionally, any open-text responses will be analysed using a general inductive approach (Thomas, 2006), with results used to provide a description of how the intervention worked in practice. In addition, this data will be used to empirically examine the assumptions underlying the key mechanisms and processes²⁴ in Triple P's logic model, exploring how they worked in the context of this evaluation. This modelling is considered a key way of understanding the association between activities and outputs to their intended outcomes, and is considered particularly important when the change in distal outcomes (in this case, language) is assumed to be underpinned in some way by direct effects on proximal outcomes (in this case, child behaviour) (Humphrey et al., 2016).

A description of the counterfactual will also be provided based on the responses from practitioner surveys in control settings.

²² We intend to send texts on five SDQ items. The items will be selected in consultation with the Triple P team.

²³ In our original proposal we suggested a baseline practitioner survey, in addition to the end-line practitioner survey, in order to collect an understanding of usual practice. However, upon consideration, we felt that the available resources would be better used to conduct parent interviews. We feel that the fact that children will be new to the setting coupled with the fact that eligibility criteria is that settings have no prior experience of delivering an evidence-based parenting programme, means that usual practice is less likely to have a significant impact on how Triple P operates in practice. That said, we will be sure to include questions in the end-line survey in both arms to collect information on how settings i) work with parents, ii) work with parents to improve behaviour and language in the home, and iii) details of how these activities may or may not have differed in the evaluation year compared to previous years.

²⁴ In this context key mechanisms relate to change in parenting skills. It is assumed that the programme will positively impact the parenting style and how well parents learn to cope with a broad range of behavioural problems.

Practitioners will also be asked to provide information on costs associated with the delivery of Triple P (in treatment arms) or similar parenting/home language programmes (in control arms, if applicable).²⁵ Questions will be designed to ensure adherence to the EEF's guidance and will collect data on both financial costs and practitioner time.

Data on costs will be presented descriptively in the relevant section of the EEF report template.

6. PHONE INTERVIEWS WITH PARENTS

We propose conducting phone interviews with parents from the treatment settings, as phone interviews have the benefit of being cost effective and minimise time burden on parents. Telephone interviews have been used successfully with parents in other EEF trials, including the Family Skills evaluation (Husain, et al. 2018).

We propose using a stratified sampling (see McDavid et al., 2019) to draw views from a meaningful sample of parents in the treatment settings. To ensure that we have cases from strategically important groups, we will first rank the settings according to frequency of parent attendance (i.e. high, medium, low), using the parent attendance logs. We believe that parent attendance at sessions will be a strong indicator of how well the programme is operating in practice. Two settings from each ranking will be chosen. Once the settings have been chosen, practitioners will be asked to identify parents to approach for interview. Practitioners will be asked to identify parents from across a range of different family incomes and from different ethnicities, as these characteristics have been associated with receptive language outcomes in previous studies (Pungello, et al. 2009). The evaluation team will seek to collect views from fifteen parents with the knowledge that it is unlikely that we will be able to achieve a truly representative sample. We also propose using a £15 high street voucher as an incentive for each parent, as this has been used successfully for parent telephone interviews in previous trials (see Hussain, et al. 2018).

Parent interviews will focus on examining the intervention mechanisms of change, particularly any changes to the home learning environment and child language development in relation to parenting strategies. In designing an interview protocol, we will consult examples of other studies in the parenting field of qualitative studies oriented towards examining intervention mechanisms of change: e.g. Giusto et al., 2017, Doubt et al., 2017. The interviews will also examine parent preferences for programme delivery (Sonuga-Barke et al., 2018).

A general inductive approach will be used to identify relevant themes and/or categories most relevant to the research objectives (Thomas, 2006). A description of the most important themes will be presented in the final report. This will also allow us to further our understanding of how the logic model worked in practice and the conditions needed to make Triple P succeed.

²⁵ In the original proposal we discussed having this completed by head teachers, however, we think that it would be more cost-effective to ask practitioners to respond to this, with a caveat that they can/should ask their heads if they cannot answer the question.

Cost evaluation

Cost data will be gathered through practitioner online surveys in the implementation and process evaluation. Questions will be targeted at assessing any pre-requisite costs (such as training costs and materials) and any direct and marginal costs directly attributable to early years settings participation in the intervention (printing, staff time, cover, etc.). We will use this information to estimate cost per-child, following EEF guidelines (EEF, 2015).

The main costs of the intervention relate to training, materials, and the time of practitioners and other setting staff to complete the programme activities. To calculate the cost of training and materials, the evaluation team will rely on data provided by the delivery team. RAND Europe will also take into account the cost of the time of practitioners and other staff in delivering the programme.

Questions will be targeted at assessing any pre-requisite costs (such as training costs and any necessary materials) and any direct and marginal costs directly attributable to the early years settings participation in the intervention (printing, staff time, cover, etc.).

We will use the information on direct and indirect costs to estimate cost per-child, following EEF guidelines (EEF, 2018).

Data protection

(Please see EEF's GDPR Briefing for Evaluators (March 2018) [here](#).)

The trial has been registered on the ISRCTN registry, which stands for 'International Standard Randomised Controlled Trial Number' and is used to describe RCTs and efficacy trials at inception. The trial has been assigned an ID registration number: ISRCTN89357177.

The ethics and registration processes are in accordance with the ethics policies adopted by RAND Europe. The evaluation is currently reviewed by RAND U.S. Human Subjects Protection Committee (HSPC).

Prior to baseline data for the children being sent to the evaluation team, parents will be sent information and withdrawal forms by the early years settings and have the opportunity to return these. The parental information sheets and withdrawal forms will be sent out to parents by the early years settings after the school representative sign the Memorandum of Understanding (MoU) describing what is involved in the trial. Parents can withdraw their children at any time from the research, but the initial withdrawal forms can be returned by parents within two weeks.

If participants choose to withdraw their children from the study later on, their data will not be collected or will be deleted, as appropriate. RAND Europe will collect consent forms for any parents who agree to participate in an interview. Furthermore, the cover page for each survey will contain a privacy notice for respondents. It will inform respondents that participation in the survey is entirely voluntary.

None of the evaluation team has any conflicts of interest and all members of the study team have approved this protocol prior to publication.

The lawful basis for RAND Europe (the data controller) processing the data under the GDPR is their 'legitimate interest'. That is, they have a legitimate interest in processing the data in order to identify settings that meet the recruitment criteria, to approach those settings and to work with the settings that wish to participate in the trial. In considering whether they could rely on legitimate interests as the lawful basis for processing the data, the data controllers have balanced their interests with the interests of the data subjects. The data subjects' data will not be used in any way that is detrimental to their rights and/or freedoms. On this basis the data controllers have assessed a legal basis of legitimate interests to be applicable. RAND Europe shall obtain

baseline and outcome data from the early years settings, as well as its subcontractor (e.g., Elklan), who will act as a processor pursuant to appropriate data sharing terms in its subcontract. Data obtained by Elklan will be obtained on the basis of legitimate interests and children and parents shall be provided with age-appropriate fair processing privacy notices that explain the use, storage and secure handling of the data.

Data will be shared securely using specialised encrypted software (e.g. Syncplicity or Sharepoint for Research).

Data will only be saved on General Data Protection Regulation (GDPR) compliant, secure servers inside the EEA or UK. All processes will be handled in accordance with RAND's Data Protection Policy. RAND Europe is registered with Information Commissioner's Office (ICO), registration number Z6947026 and is certified for adhering to ISO 9001:2015 quality management practices.

For the purpose of research, following the completion of the trial, the data will be shared with the EEF archive, at which point EEF will act as the data controller.

Personnel

DELIVERY TEAM: TRIPLE P UK

Overall Project Lead and Manager and Implementation Consultant: Dr Claire Halsey

Triple P UK Chief Executive Officer: Mr Matt Buttery

EVALUATION TEAM: RAND EUROPE

Overall Project & Evaluation Lead: Elena Rosa Brown (RAND Europe)

Project Manager: Dr. Sashka Dimova (RAND Europe)

Core fieldwork and analysis team: Dr. Sashka Dimova (RAND Europe) | Amelia Harshfield (RAND Europe) | Natalie Picken (RAND Europe) | Andreas Culora (RAND Europe)

Risks

Risk	Assessment	Mitigation strategy
Recruitment failure	Likelihood: Medium Impact: High	There were challenges in recruiting a sufficient number of school based settings. As a result the number of targeted settings was widened in May, 2019. The delivery and evaluation team are in dialogue over any recruitment issues. Furthermore, the EEF provides support to encourage recruitment.
Selection of children	Likelihood: Low Impact: High	The reason for inclusion in the trial will be recorded by practitioners.
Attrition	Likelihood: Moderate Impact: Moderate to high	There's risk of parent attrition due to typically long delay between parent recruitment and group delivery. Mitigation provision of retention support materials to intervention settings. The study is powered to detect an effect of d=0.20 or smaller even if there are 130 settings involved. We proposed recruiting

		150 settings to build in a 'buffer' for attrition at person and setting level. Settings and parents are given clear information about participation before signing up. Attrition to be monitored and reported according to CONSORT guidelines (Campbell et al., 2010).
Different rates of attrition from control and intervention groups	Likelihood: Low Impact: Moderate	There is a risk that early years settings in the intervention group may face an extra burden in terms of time and resources to deliver the programme. This can be mitigated by regular liaison with settings to secure continued engagement in the trial. Early years settings would have agreed to the terms of the MoUs, which include the commitment for data to be collected at various stages.
Missing data	Likelihood: Moderate Impact: Moderate	To limit the amount of missing data screening/testing will happen in an extended period (approximately one month). In the event of more than 10% missing data then the strategies set out above will be followed.
Children's mobility	Likelihood: Moderate Impact: Moderate	As the sample of targeted settings is broader, including private settings, the risk of children dropping out due to children moving settings is significant. By limiting the proportion of private settings to maximum of 20% we will mitigate the risk of significant drop out rates.
Low implementation fidelity	Likelihood: Low to moderate Impact: Moderate	Process evaluation to monitor and document fidelity of implementation. Remain in dialogue with the Triple P Delivery Team regarding their view on fidelity and on findings solutions.
Cross-contamination	Likelihood: Moderate Impact: High	Clear instructions will be provided to participants about the trial to avoid contamination.
Evaluation team members absence or turn-over	Likelihood: Moderate Impact: Low	All RAND Europe staff have a three month notice period to allow sufficient time for handover. The team can be supplemented by researchers with experience in evaluation from the larger RAND Europe pool.
Low response rates for online surveys	Likelihood: Moderate Impact: Moderate	Online surveys to be kept to a maximum of 5-15 minutes long. Respondents given the opportunity to complete survey online on multiple occasions if required. Sufficient data collection window given with real-time monitoring of response rates to allow for reminders to be targeted.
Lack of coordination with larger teams	Likelihood: Moderate Impact: Moderate	Teams to attend initial meetings and agree on roles and responsibilities at the outset. Regular updates to be provided to the lead evaluators. Regular contact between senior team from each organisation.

Timeline

Dates	Activity	Staff responsible/leading
April-August, 2019	Recruitment	Triple P UK
July, 2019	Practitioner/setting selection	Triple P UK
September, 2019	Parent recruitment	Triple P UK /Settings
September, 2019	Post paper SDQ T to all settings	RAND Europe with contact info from Triple P UK

September-October	Complete SDQ T online scoring using RAND survey platform, write down the total problem scores	Setting practitioners
4 th October	Teacher SDQ, children and parent Information (baseline data) completed and shared with RAND	Setting practitioners
September-October (by October 9 th)	Using SDQ Ts and other criteria, select eligible children and communicate with their parents to select the parenting group	Setting practitioners
9-20 October	Randomisation occurs	RAND Europe
23 rd October	Implementation Planning Consultation begins with Intervention Sites.	Triple P UK /Settings
Mid October 2019	Practitioner Training resources arrive to venues	Triple P UK
4 th - 8 th November	Setting Practitioner Briefings	Triple P UK
Mid November	Practitioner Level 4 Group Triple P Training	Triple P UK
January 2020	Practitioner Pre-accreditation and Accreditation	Triple P UK /Setting
January 27 th week	Clinical workshops	Triple P UK / Setting
February 3 rd week (break for half term 10 th April 2020)	Group Triple P delivery Collection of parental attendance logs	Setting
April- June 2020	Online surveys with practitioners and parents Interviews with parents SDQ text messages to parents in control settings	RAND Europe
May 31 st 2020	Parent Delivery completed Parent pre- and post-data scored centrally / shared with RAND	Setting
June- July, 2020	SDQ-T survey completed by early years practitioners	RAND Europe/Settings
22 nd June- 12 July, 2020	Post-trial primary outcome testing	RAND Europe/ Elklan
August-October, 2020	Analysis	RAND Europe
November-December, 2020	Final report	RAND Europe

Appendix A

DESCRIPTION OF WHAT IS COVERED IN EACH SESSION WITH PARENTS

Session 1: This session provides parents with an introduction to positive parenting, why children behave as they do, and how to set goals for change.

Session 2: In the second group session, the practitioner discusses how to develop good relationships with children, how to encourage good behaviour, and the four strategies for how parents can teach their children new skills and behaviours.

Session 3: During the third group session, the practitioner offers additional strategies to assist parents with managing misbehaviour during this session. Parents will also learn to develop parenting routines to promote compliance and manage non-compliance from their children. They have an opportunity to rehearse these routines during the session.

Session 4: This session covers family survival tips, identifying high risk situations that still cause concern, and how to develop planning ahead routines to promote good child behaviour in high risk situations (e.g. shopping, learning how to take turns, fighting with siblings, getting ready for school). Parents also prepare for their individual consultations during this session.

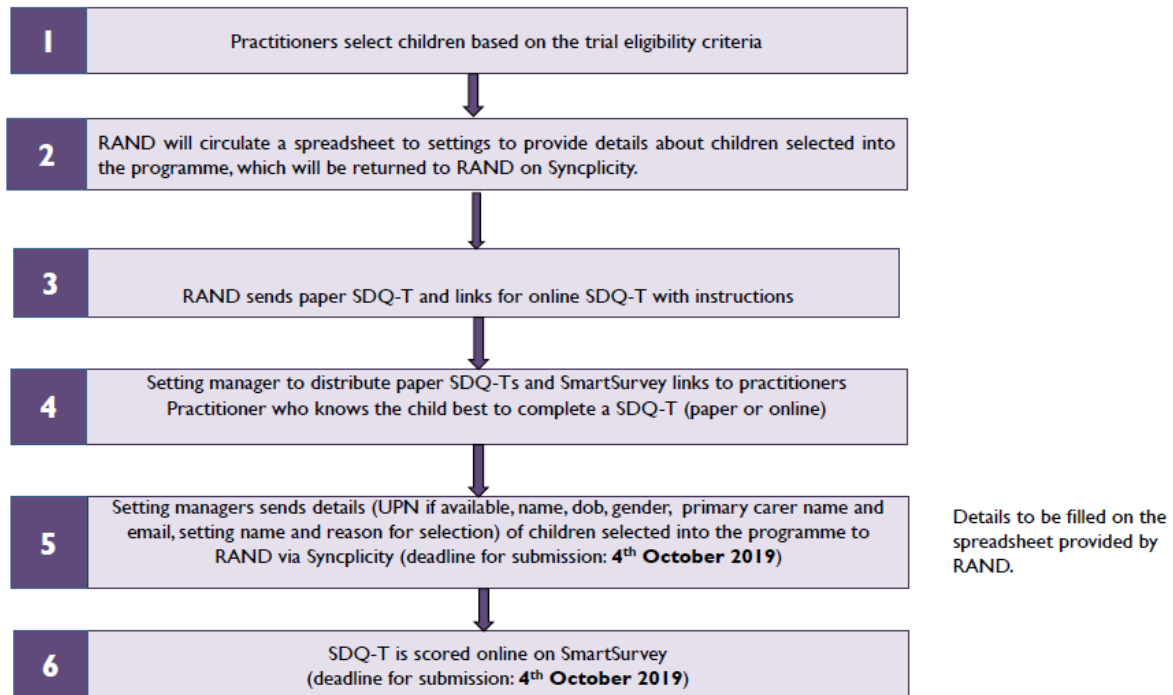
Session 5-7: The practitioner provides feedback from initial assessments that the family completed and then uses the self-regulatory feedback model to help parents review their implementation of planning ahead routines for their high-risk situations. From this, parents set goals for further refinement of their routines, if needed.

Session 8: Parents return for a final group session to review progress, look at ways to maintain changes and plan for the future, and to close the program. If necessary, referral options are discussed.

Appendix B

DESCRIPTION OF BASELINE DATA COLLECTION PROCESS

Figure 2: SDQ T data collection process



References

- Allison, P. D., (2012) Handling Missing Data by Maximum Likelihood. Haverford, PA: Statistical Horizons. Retrieved from: <https://statisticalhorizons.com/wp-content/uploads/MissingDataByML.pdf>
- Allen, G. (2011). Early intervention: the next steps, an independent report to Her Majesty's government by Graham Allen MP. The Stationery Office.
- Asmussen, K., & Weizel, K. (2010). Evaluating the evidence: Fathers, families and children. National Academy for Parenting Practitioners.
- Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological review*, 84(2), 191.
- Bandura, A. (Ed.). (1995). Self-efficacy in changing societies. Cambridge university press.
- Bröning, S., Baldus, C., Thomsen, M., Sack, P. M., Arnaud, N., & Thomasius, R. (2017). Children with Elevated Psychosocial Risk Load Benefit Most from a Family-Based Preventive Intervention: Exploratory Differential Analyses from the German “Strengthening Families Program 10–14” Adaptation Trial. *Prevention Science*, 18(8), 932-942.
- Carril, A. (2017). Dealing with misfits in random treatment assignment. *Stata Journal*, 17(3), 652-667.
- Charman, T., Ricketts, J., Dockrell, J. E., Lindsay, G., & Palikara, O. (2015). Emotional and behavioural problems in children with language impairments and children with autism spectrum disorders. *International Journal of Language & Communication Disorders*, 50(1), 84-93.
- Chung, S., Leung, C., & Sanders, M. (2015). The Triple P–Positive Parenting Programme: the effectiveness of group Triple P and brief parent discussion group in school settings in Hong Kong. *Journal of children's services*, 10(4), 339-352.
- Coyne, J. C., & Kwakkenbos, L. (2013). Triple P-Positive Parenting programs: the folly of basing social policy on underpowered flawed studies. *BMC medicine*, 11(1), 11.
- Crisante, L., & Ng, S. (2003). Implementation and process issues in using Group Triple P with Chinese parents: Preliminary findings. *Australian e-Journal for the Advancement of Mental Health*, 2(3), 226-235.
- Croft, S., Stride, C., Maughan, B., & Rowe, R. (2015). Validity of the strengths and difficulties questionnaire in preschool-aged children. *Pediatrics*, 135(5), e1210-e1219.
- De Graaf, I., Speetjens, P., Smit, F., de Wolff, M., & Tavecchio, L. (2008). Effectiveness of the Triple P Positive Parenting Program on behavioral problems in children: A meta-analysis. *Behavior Modification*, 32(5), 714-735.
- Denham, S. A., Workman, E., Cole, P. M., Weissbrod, C., Kendziora, K. T. & Zahn-Waxler, C., (2000). Prediction of externalizing behaviour problems from early to middle childhood: the role of parental socialisation and emotion expression. *Development and Psychopathology*, 12, pp. 23–45.
- Dong, N., & Maynard, R. (2013). PowerUp!: A tool for calculating minimum detectable effect sizes and minimum required sample sizes for experimental and quasi-experimental design studies. *Journal of Research on Educational Effectiveness*, 6(1), 24-67.
- Doyle, O., Hegarty, M., & Owens, C. (2018). Population-based system of parenting support to reduce the prevalence of child social, emotional, and behavioural problems: Difference-in-differences study. *Prevention Science*, 19(6), 772-781.

Early Intervention Foundation (2016) Foundations for life: what works to support parent child interaction in the early years. London: Early Intervention Foundation. Retrieved from <https://www.eif.org.uk/files/pdf/foundations-for-life.pdf>

Education Endowment Foundation (2015) EEF Guidance on Cost Evaluations. London: Education Endowment Foundation. Retrieved from: https://v1.educationendowmentfoundation.org.uk/uploads/pdf/EEF_guidance_to_evaluators_on_cost_evaluation.pdf

Education Endowment Foundation (2018) Statistical Guidance for EEF evaluations. London: Education Endowment Foundation. Retrieved from: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol_or_SAP/EEF_statistical_analysis_guidance_2018.pdf

Ezpeleta, L., Granero, R., de la Osa, N., Penelo, E., & Domènech, J. M. (2013). Psychometric properties of the Strengths and Difficulties Questionnaire3–4 in 3-year-old preschoolers. *Comprehensive Psychiatry*, 54(3), 282–291.

Gardner, F., Sonuga-Barke, E. & Sayal, K., (1999). Parents anticipating misbehaviour: an observational study of strategies parents use to prevent conflict with behaviour problem children. *Journal of Child Psychology and Psychiatry*, 40, pp. 1185–96.

Gardner, F., Leijten, P., Mann, J., Landau, S., Harris, V., Beecham, J., et al. (2017). Could scale-up of parenting programmes improve child disruptive behaviour and reduce social inequalities? Using individual participant data meta-analysis to establish for whom programmes are effective and cost-effective. *Public Health Research*, 5(10). Available online: https://www.ncbi.nlm.nih.gov/books/NBK469806/pdf/Bookshelf_NBK469806.pdf

Gibson, T. A., Peña, E. D., & Bedore, L. M. (2014). The relation between language experience and receptive-expressive semantic gaps in bilingual children. *International Journal of Bilingual Education and Bilingualism*, 17(1), 90–110.

Glennerster, R., & Takavarasha, K. (2013). *Running randomized evaluations: A practical guide*. Princeton University Press.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry*, 38 (5), 581–86.

Gould, Nick, and Joanna Richardson. "Parent-training/education programmes in the management of children with conduct disorders: Developing an integrated evidence-based perspective for health and social care." *Journal of Children's Services* 1.4 (2006): 47–60.

Hahlweg, K., Heinrichs, N., Kuschel, A., Bertram, H., & Naumann, S. (2010). Long-term outcome of a randomized controlled universal prevention trial through a positive parenting program: is it worth the effort?. *Child and adolescent psychiatry and mental health*, 4(1), 14.

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

Heß, S. (2017). Randomization inference with Stata: A guide and software. *The Stata Journal*, 17(3), 630–651.

Higgins, J. P. T., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., Savović, J., Schulz, K. F., Weeks, L., & Sterne, J. A. C. (2011) The Cochrane Collaboration's tool for assessing risk of bias in randomised trials, *British Medical Journal*, 18(343) (d5928)

Higgins, J. P. T., Savović, J., & Page, M. J. (2016). Revised Cochrane risk of bias tool for randomized trials (RoB 2.0). Available at: <https://sites.google.com/site/riskofbiastool/welcome/rob-2-0-tool>

- Hoath, F. E., & Sanders, M. R. (2002). A feasibility study of Enhanced Group Triple P—positive parenting program for parents of children with attention-deficit/hyperactivity disorder. *Behaviour Change*, 19(4), 191-206.
- Hox, J. (1998). Multilevel modeling: When and why. In Balderjahn I., Mathar R., Schader M. (eds) *Classification, Data Analysis, and Data Highways. Studies in Classification, Data Analysis, and Knowledge Organization*. Berlin, Heidelberg: Springer.
- Husain, F., Wishart, R., Marshall, L., Frankenberg, S., Bussard, L., Chidley, S., Hudson, R., Votjkova, M. & Morris, S. 2018. Family Skills Evaluation Report. Education Endowment Foundation.
- Hutchings, J., Bywater, T. & Daley, D., (2007). Early prevention of conduct disorder: how and why did the North and Mid Wales Sure Start study work?. *Journal of Children's Services*, 2(2), pp. 4–15.
- Hutchison, D., & Styles, B. (2010). *A guide to running randomised controlled trials for educational researchers*. Slough: NFER.
- Kim, J. H., Schulz, W., Zimmerman, T., & Hahlweg, K. (2018). Parent-Child Interactions and Child Outcomes: Evidence from Randomized Intervention. *Labour Economics*.
- Lindsay, G., Strand, S., & Davis, H. (2011). A comparison of the effectiveness of three parenting programmes in improving parenting skills, parent mental-well being and children's behaviour when implemented on a large scale in community settings in 18 English local authorities: the parenting early intervention pathfinder (PEIP). *BMC public health*, 11(1), 962.
- Little, M., Berry, V. L., Morpeth, L., Blower, S., Axford, N., Taylor, R., Bywater, T., Lehtonen, M. and Tobin, K. (2012). The impact of three evidence based programmes delivered in public systems in Birmingham, UK. *International Journal of Conflict and Violence*, 6 (2), 260-272
- Malti, T., Ribeaud, D., & Eisner, M. P. (2011). The effectiveness of two universal preventive interventions in reducing children's externalizing behavior: a cluster randomized controlled trial. *Journal of Clinical Child & Adolescent Psychology*, 40(5), 677-692.
- Marryat, L., Thompson, L., & Wilson, P. (2017). No evidence of whole population mental health impact of the Triple-P parenting programme: findings from a routine dataset. *BMC Pediatrics*, 17(1), 40.
- McDavid, J. C., Huse, I. & Hawthorn, L.R.L. (2019). *Program Evaluation and Performance Measurement*. Thousand Oaks: Sage.
- Mol, S. E., Bus, A. G., & De Jong, M. T. (2009). Interactive book reading in early education: A tool to stimulate print knowledge as well as oral language. *Review of Educational Research*, 79(2), 979-1007.
- Nowak, C., & Heinrichs, N. (2008). A comprehensive meta-analysis of Triple P-Positive Parenting Program using hierarchical linear modeling: Effectiveness and moderating variables. *Clinical child and family psychology review*, 11(3), 114.
- Oakes, J. M. (2013) Effect identification in comparative effectiveness research. *The Journal for Electronic Health Data and Methods*, 1(1):1004. doi: 10.13063/2327-9214.1004
- Patterson, G. R., Chamberlain, P., & Reid, J. B. (1982). A comparative evaluation of a parent-training program. *Behavior therapy*, 13(5), 638-650.
- Payne, A. C., Whitehurst, G. J., & Angell, A. L. (1994). The role of home literacy environment in the development of language ability in preschool children from low-income families. *Early Childhood Research Quarterly*, 9(3-4), 427-440.

- Prinz, R. J., Sanders, M. R., Shapiro, C. J., Whitaker, D. J., & Lutzker, J. R. (2009). Population-based prevention of child maltreatment: The US Triple P system population trial. *Prevention science*, 10(1), 1-12.
- Pungello, E. P., Iruka, I. P., Dotterer, A. M., Mills-Koonce, W. R., & Reznick, J. S. (2009). The effects of income, race, and parenting practices on language development in early childhood. *Developmental Psychology*, 45(2), 544-557.
- Robinson-Smith, L., Merrell, C., Stothard, S., Torgerson, C. (2018) Independent Evaluation of the EasyPeasy programme. Available at: <https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/easypeasy-learning-through-play/>
- Rogde, K., Melby-Lervåg, M., & Lervåg, A. (2016). Improving the general language skills of second-language learners in kindergarten: A randomized controlled trial. *Journal of Research on Educational Effectiveness*, 9(sup1), 150-170.
- Rutter, M. (1985). Resilience in the face of adversity: Protective factors and resistance to psychiatric disorder. *The British Journal of Psychiatry*, 147(6), 598-611.
- Rutter, M., Giller, H. & Hagell, A., (1998). *Antisocial Behaviour in Young People*. Cambridge: Cambridge University Press.
- Sanders, M. R. (1999). The Triple P-Positive parenting program: Towards an empirically validated multilevel parenting and family support strategy for the prevention of behavior and emotional problems in children. *Clinical Child and Family Psychology Review*, 2(2), 71–90.
- Sanders, M. R., Cann, W., & Markie-Dadds, C. (2003). The Triple P-Positive Parenting Programme: a universal population-level approach to the prevention of child abuse. *Child Abuse Review: Journal of the British Association for the Study and Prevention of Child Abuse and Neglect*, 12(3), 155-171.
- Sanders, M. R. (2008). Triple P-Positive Parenting Program as a public health approach to strengthening parenting. *Journal of family psychology*, 22(4), 506.
- Sanders, M. R., Kirby, J. N., Tellegen, C. L., & Day, J. J. (2014). The Triple P-Positive Parenting Program: A systematic review and meta-analysis of a multi-level system of parenting support. *Clinical psychology review*, 34(4), 337-357.
- Schochet, P. Z., & Chiang, H. S. (2011). Estimation and identification of the complier average causal effect parameter in education RCTs. *Journal of Educational and Behavioral Statistics*, 36(3), 307-345.
- Scott, S. (2010). National dissemination of effective parenting programmes to improve child outcomes. *The British Journal of Psychiatry*, 196(1), 1-3.
- Senn, S. (2004). Individual response to treatment: is it a valid assumption?. *Bmj*, 329(7472), 966-968.
- Sim, M., Bélanger, J., Hocking, L., Dimova, S., Iakovidou, E., Janta, B., & EIF, W. T. (2018). *TEACHING, PEDAGOGY AND PRACTICE IN EARLY YEARS CHILDCARE: AN EVIDENCE REVIEW ANNEXES TO THE REPORT*.
- Smith, G. C., Hayslip Jr, B., Hancock, G. R., Strieder, F. H., & Montoro-Rodriguez, J. (2018). A randomized clinical trial of interventions for improving well-being in custodial grandfamilies. *Journal of Family Psychology*, 32(6), 816.
- Snijders, T. A. B., & Bosker, R. J. (2012). Discrete dependent variables. *Multilevel analysis: an introduction to basic and advanced multilevel modeling*, 304-307.
- Snijders, T. A. (2005). Power and sample size in multilevel modeling. *Encyclopedia of statistics in behavioral science*, 3, 1570-1573.

- Snijders, T. A., & Bosker, R. J. (1994). Modeled variance in two-level models. *Sociological methods & research*, 22 (3), 342-363.
- Sonuga-Barke, E. J., Barton, J., Daley, D., Hutchings, J., Maishman, T., Raftery, J., et al (2018). A comparison of the clinical effectiveness and cost of specialized individually delivered parent training for preschool attention deficit/hyperactivity disorder and a generic, group based programme: a multi centre, randomised controlled trial of the New Forest Parenting Programme versus Incredible Years. *European Child and Adolescent Psychiatry*, 27(6), 797-809. DOI: 10.1007/s00787-017-1054-3
- Spybrook, J., Kelcey, B., & Dong, N. (2016). Power for detecting treatment by moderator effects in two-and three-level cluster randomized trials. *Journal of Educational and Behavioral Statistics*, 41(6), 605-627.
- Thomas, D. R. (2006). A General Inductive Approach for Analyzing Qualitative Evaluation Data. *American Journal of Evaluation*, 27(2): 237 – 246.
- Wilson, P., Rush, R., Hussey, S., Puckering, C., Sim, F., Allely, C. S., ... & Gillberg, C. (2012). How evidence-based is an'evidence-based parenting program'? A PRISMA systematic review and meta-analysis of Triple P. *BMC medicine*, 10(1), 130.
- WHO. (2016). INSPIRE Seven Strategies for Ending Violence Against Children. World Health Organization, 1–108.
- Zubrick, S. R., Ward, K. A., Silburn, S. R., Lawrence, D., Williams, A. A., Blair, E., ... & Sanders, M. R. (2005). Prevention of child behavior problems through universal implementation of a group behavioral family intervention. *Prevention Science*, 6(4), 287.