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

Changes to protocol

Due to recruitment being lower than expected in 2019/2020, a second year of recruitment was being planned for the 2020/2021 academic year. Some elements of the first year of the trial were run as planned including:

- baseline testing of children using teacher assessed SDQ-T in Autumn 2019
- baseline survey in all schools in Autumn 2019
- 67 settings being randomised and informed of their allocation in Autumn 2019
- Intervention setting practitioners trained and accredited in the Triple P programme and delivery of the programme commenced as planned

However, due to closure of early years settings to most children as a result of national lockdown from Spring to Summer 2020. Despite the reopening of settings in June, the possibility of creating a comparable group in the second year was no longer considered feasible. Owing to the minimal power associated with the given sample size and the likelihood of high attrition, the decision was taken to not collect evaluation-specific impact outcomes (i.e. SDQ-T, CELF). It should be noted that pre and post Triple P Group outcomes for the intervention settings were collected (i.e. Parenting Scale, parent-report SDQ, DASS 21) in line with programme requirements. The evaluation was rescoped to focus on data from the process evaluation (e.g. interviews, survey, aggregate data on parenting self-report measures collected by Triple P).

Table 1: Changes to evaluation

Planned	Current
150 settings randomised	67 setting randomised
Post-intervention primary and secondary outcome testing Summer 2020 (i.e. CELF, SDQ-T)	No testing of primary and secondary outcomes (i.e. CELF, SDQ-T)
Analysis of primary and secondary impact outcomes (i.e. CELF, SDQ-T)	No analysis of primary and secondary outcomes (i.e. CELF, SDQ-T)
Collect parent views of behaviour via text messages to parents in control arm	Replaced with phone interviews with parents in control arm
	Interviews with practitioners in treatment and control arms
	Increased focus on impact of Covid-19 and impact on families, settings, and implications for delivery of Triple

This protocol updates the previous protocol and includes all elements of the evaluation that were delivered as planned (i.e. baseline and randomisation) and the changes that were made subsequently.

- p. 2 –updated project overview table
- p. 7 – added additional information on how Triple P adapted the final session and subsequent data collection in light of Covid-19 restrictions
- p. 9 – added rationale for increasing IPE activities and clarification on impact evaluation activities that were forgone
- p. 10 – added overview of changes to impact evaluation
- p. 10 – updated randomisation – based on geographical location and type of nursery
- p. 15 – details of MDES size based on numbers randomised added
- p.17 – added rationale for why outcome testing was cancelled
- p. 18 – removed referencing to testing in settings with more than 10 children
- p. 19 – added rationale for why outcome testing was cancelled
- p. 19 – added background context to analysis section, removed paragraphs in analysis section that are no longer relevant given the lack of outcome testing

- p. 20 – added IPE methods overview table (Table 3)
- p. 21 – removed paragraphs in analysis section that are no longer relevant given the lack of outcome testing
- p. 22 – added Covid-19 research question and rationale for its inclusion
- p. 22 – updated the table to reflect new IPE activities
- p. 22 and 24 – removed references to analysis of parent texts
- p.23 – added summary of changes to IPE activities
- p. 24 – added details on how data from Triple P (i.e. practitioner questionnaires, parent measures) will be analysed and relevance to research questions
- p. 24 – removed references to compliance measure
- p. 25 – added more specificity on how practitioner survey data in both arms will be analysed and relevance to research questions
- p. 25 – removed footnote as repeated in text.
- p.26 – added Covid-19 related questions into parent interview and rationale for interviews with parents in control arm
- p. 26 – added more specificity on how information from parent interviews in both arms will be analysed and relevance to research questions
- p. 27 – updated sampling approach for practitioner interviews
- p. 27 – added details on updated privacy notice
- p. 32 – updated timeline to reflect changes in activities and disruptions caused by Covid-19

Originally planned:

	Original deign	**Updated design** (only changes to design noted)
PROJECT TITLE	Level 4 Group Triple P	
DEVELOPER (INSTITUTION)	*Parenting and Family Support Centre, University of Queensland*	
EVALUATOR (INSTITUTION)	RAND Europe	
PRINCIPAL INVESTIGATOR(S)	Elena Rosa Brown	
PROTOCOL AUTHOR(S)	Dr. Sashka Dimova, Elena Rosa Brown, Amelia Harshfield, Dr. Andreas Culora, Dr. Alex Sutherland, Natalie Picken	
TRIAL DESIGN	Two-arm cluster randomised controlled trial with random allocation at the Early Years setting level	Implementation and process evaluation
PUPIL AGE RANGE AND	3 and 4, Early Years (Nursery)	

KEY STAGE		
NUMBER OF SCHOOLS	150	67
NUMBER OF PUPILS	1,800	549
PRIMARY OUTCOME	Expressive language	n/a
SECONDARY OUTCOME	Child behaviour	n/a

Protocol version history

VERSION	DATE	REASON FOR REVISION
1.0	06 June, 2019	
2.0	22 September, 2020	Due to recruitment being lower than expected in 2019/2020, a second year was planned for the 2020/2021 academic year. However, due to school closures as a result of national lockdown creating a comparable group in the second year was not considered feasible. Owing to the minimal power associated with the given sample size the decision was taken to not collect impact outcomes (i.e. CELF, SDQ-T). The evaluation was rescoped to focus on data from the process evaluation (e.g. interviews, survey, aggregate data on parenting self-report measures collected by Triple P).

Trial Evaluation Protocol

Level 4 Group Triple P : Positive Parenting Program®

Evaluator (institution): RAND Europe
Principal investigator(s): Elena Rosa Brown

Education Endowment Foundation

Template last updated: March 2018

Table of contents

Trial Evaluation Protocol

Level 4 Group Triple P : Positive Parenting Program®

Evaluator (institution): RAND Europe
Principal investigator(s): Elena Rosa Brown

Education Endowment Foundation

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 et al., 1982), child and family behaviour therapy and applied behaviour analysis (Portfield, et al. 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 et al., 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 4year 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 practitioners[1] 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.

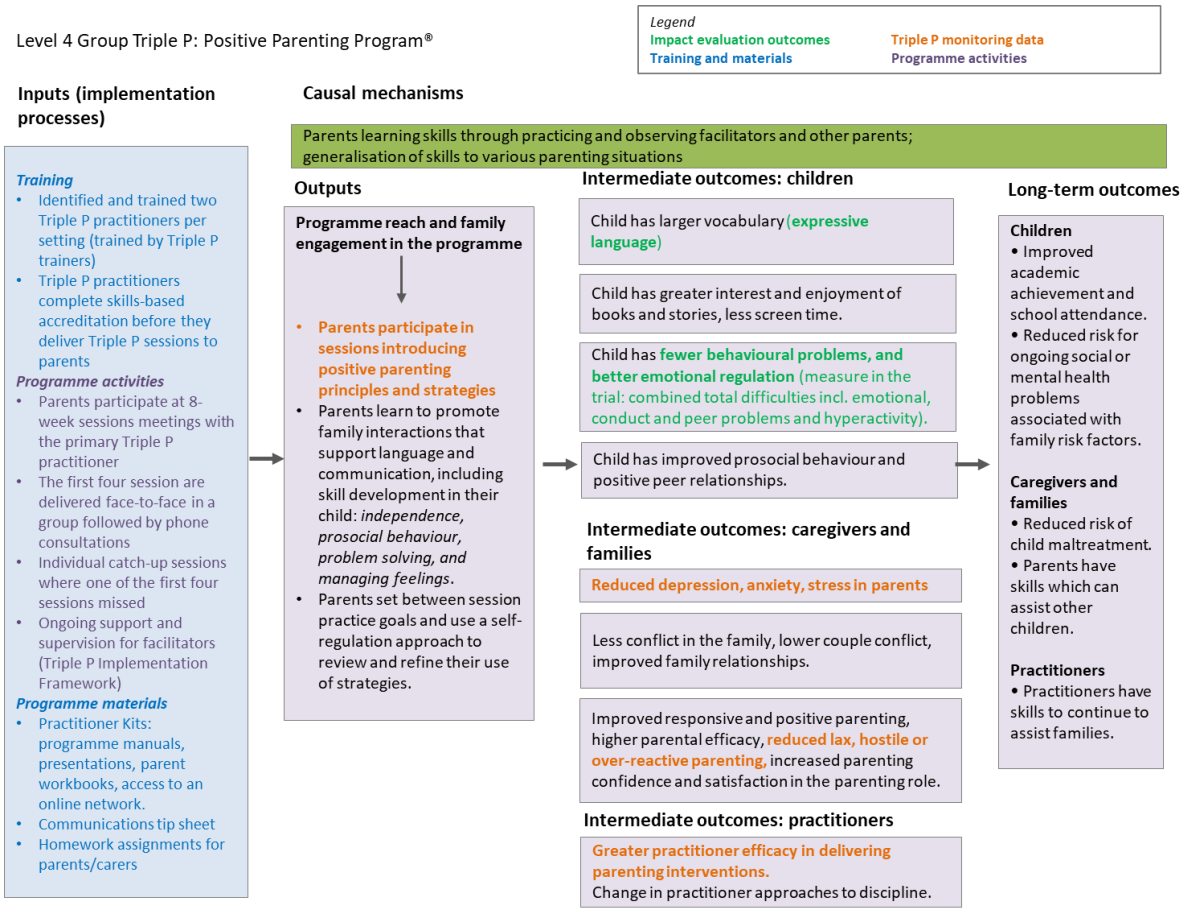
[1] Training is offered to two practitioners per setting in order to mitigate risk of attrition. However, it is not mandatory criteria for settings to have two practitioners at the training.
[2] Triple P trainers are trained facilitators, experienced and accredited contractors or members of TP staff, typically holding psychology qualifications (usually MSc or CPsychol).

  - 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.
  - 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



Update: Adaptations due to Covid-19

To support practitioners in delivering Session 8 and collecting post-assessment data during lockdown the following was provided by Triple P UK:

- Phone consultations took place with a small number of intervention setting practitioners to gauge the feasibility of delivering Session 8 and collecting the post course Triple P outcome measures (i.e. those routinely collected as part of programme implementation). This consultation concluded that the appetite and potential to complete the programme was there, and as such the decision to proceed was agreed with the EEF and RAND.
- Practitioners were provided with guidance by email to support remote delivery of Session 8, and remote collection/return of post -group parent assessments, including a step by step guide to collecting outcomes measures by phone.
- Email and phone support was offered to all intervention settings to problem solve both delivery and data collection and return.

- Intervention setting practitioners were provided with step by step guidance regarding completion of Triple P Session 8 and collection of post-group parent outcome measures in an email on 21st May 2020 and an email clearly describing the process – see below.

AREAS OF CHANGE IN DELIVERY PROTOCOL

TIMEFRAMES

The original protocol time frame between Triple P Group session 7 and Group session 8 and collection of post group measures was one week.

- Actual delay between delivery of session 7 and session 8 was 12 weeks or more.
  The 8 week Group Triple P delivery within the EEF trial started first week in February 2020, with weekly sessions delivered (excluding half term). This resulted in many/most settings reaching session 7 (of 8) prior to COVID 19 lockdown.
- The time of collection of Parent Post-Group outcome measures (collected as part of Triple P's process) was also delayed by at least 12 weeks post-delivery of session 7 as collection was due as the closing activity in Group Session 8.

DELIVERY METHOD

Original protocol delivery method of Group Triple P was face to face group sessions 1,2,3,4,8, and individual phone support session, 5,6,7.

- The final Triple P Group session 8 could not be delivered face to face following COVID 19 lockdown and school closure. This session was therefore changed to remote delivery by phone or video-conference which represents a change to protocol.
- Session 8 was delivered in the 1:1 format, rather than as a Group. This represents a change to protocol.
- Collection of Post-Group measures could not be delivered face to face as per protocol and was achieved remotely by phone or video conference with the Group Measures Booklet being posted to parents' home address for self-directed completion and stamped addresses envelop to return to setting.
- Increased time was required to complete Session 8 by delivering one to one compared to group delivery.
- Increased time was required to collect Post Group Measures through the revised process and need to collect outcome measures one to one compared to in a group session.

It should be noted that the delivery of Triple P by phone or remotely is an existing option for Triple P delivery and its effectiveness supported by research (Cann, et al., 2003) and the change in delivery is not considered by the programme team to represent a change to programme fidelity.

Study rationale and background

Recent reviews of early intervention, such as those by Allen (2011) 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 & Richardson, 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, 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 & Ng, 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 levels[3] , 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 maltreatment[4] (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.

[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.

To our knowledge, this was to 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.

Update: Due to low recruitment in the 2019/2020 academic year and closure of early years settings as a result of the national lockdown (see Changes to protocol) the trial design and impact evaluation were forgone and the implementation and process evaluation were rescoped to gather more information from practitioners and parents (see Implementation and Process Evaluation section). Research on the impact of Covid-19 has found that parents with children in the early years report a particularly negative impact on their child's social and emotional development and wellbeing, as well as additional stress for parents and carers (Pascal et al., 2020). It was agreed that an additional focus for the IPE should be the impact of Covid on families, settings, and how it has moderated the delivery and impact of Triple P.

The current evaluation was funded by the Education Endowment Foundation (EEF), Department for Education (DfE) and SHINE Trust. The programme was delivered by the Triple P UK team and the trial was carried out by RAND Europe, who are independent evaluators appointed by the EEF.

Impact Evaluation

Research questions

The core research questions this project originally sought to answer were:

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

Update: Owing to changes outlined above, these research questions have been forgone. However, details on the design of the trial are included as randomisation was completed in Autumn 2019 for the first cohort. Further details of the research that was carried out can be found in the Implementation and Process evaluation section below.

Design

Originally planned:

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)		**Geographical location**: Cluster 1: Newcastle 1; Cluster 2: Wakefield; Cluster 3: Manchester ; ; Cluster 4: Newcastle 2; Cluster 5: Preston; Cluster 6: Liverpool. [5] Type of nursery: School-based vs. private, voluntary and independent nurseries (PVIs)
Primary outcome	variable	Expressive language

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

	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 was 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 randomised within the different geographical areas and the type of nursery (school-based vs. private, voluntary and independent nurseries (PVIs)); Doing so serves to balance study arms on geographical location and by type of nursery.

Update: Due to recruitment being lower than expected in 2019/2020 (i.e. 67 schools available for randomisation, as opposed to 150 as planned), a second year of recruitment was being planned for the 2020/2021 academic year. The second year was to be run as the first year to ensure that the results could be aggregated across the two academic years. However, given school and setting closures in March 2020 and extended lockdown, the possibility of creating a comparable group in the second year was no longer considered feasible and the experimental element of the evaluation was forgone. Thus, the primary and secondary outcomes listed in the table above were not collected.

Randomisation

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

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

Settings were 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.

[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.

The recruited settings were initially organised in seven different clusters depending on the location for the practitioner training. The following training clusters had 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. However, due to lower number of settings cluster Manchester 2 was dropped.

Table 2. 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: Newcastle 2	Northumberland, Cumbria, County Durham, South Tyneside
Cluster 5: Preston	Lancashire, Blackpool
Cluster 6: Liverpool	Liverpool, Lancashire, Cheshire, Greater Manchester.

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

Baseline equivalence were examined based on the initial randomisation and results will be presented in the final report. We expected a well-conducted randomisation which would yield groups that were equivalent at baseline (Glennerster & Takavarasha, 2013). Because settings were randomly allocated to the control and intervention conditions, any imbalance at baseline will have occurred by chance. To assess imbalance at baseline, we compared 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 could take part in the Triple P programme, but up to ten families would be included in the evaluation. If a setting had more than ten participating families, then RAND Europe would have randomly selected ten families in each setting to take part in the evaluation. These families would have been selected from the lists of eligible families that were provided by the settings at the beginning of the study.[9] If a control group setting recruited twelve families, then the two "extra" families would not have been in the study and the evaluation. If an intervention group setting recruited twelve families, then the two "extra" families would be part of the intervention but were not expected to complete outcome testing.

Update: There were no settings with more than 10 families.

[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

Participants

SETTINGS

The Triple P team aimed to recruit early years settings that were school-based nurseries in the first instance. However the terms of the original protocol also outlined that they could consider recruiting non-school-based nurseries (i.e. private, voluntary and independent nurseries (PVI) or standalone state nurseries) into the trial if they found recruiting schools a challenge. As there was a risk that non-school-based nurseries would be harder to follow up and would have higher attrition, it was agreed that the number of such settings would be limited in the trial to 20% of the total cohort, ideally spread across different areas. In reality the recruitment was mixed, and priority was given to recruiting an appropriate number of settings. Further details on recruitment will be provided in the final report.

The following eligibility criteria was used for recruitment for all settings:

- The setting is located in the areas for recruitment determined by the EEF and the Delivery team[10]
- The setting has not had Triple P or Incredible Years[11] 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 was given to settings located in areas of high disadvantage.[12] Participating early years settings signed a Memorandum of Understanding (MoU) which outlined the roles and responsibilities of all stakeholders

[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 Ares (LSOA) with higher than average levels of deprivation compared to the rest of the region as measured by the Index of Multiple Deprivation (IMD).

involved. Settings and participating families were asked to consent to sharing the specified data with the delivery and evaluation team. Settings were asked to notify the delivery and/or research team immediately if a practitioner or family withdraws from data collection.

Children

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

- They were 3 or 4 years old; AND
- The teacher was concerned about a delay in a child's language / communication development. OR
- The teacher or parent was concerned about a child's behavioural, emotional or social development[13]; OR
- The practitioners decided that there was an indicator of increased family vulnerability, such as: family social disadvantage, financial stress, housing insecurity, adverse life events.

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

Parents/Carers

Parents (or carers) of nominated children were considered eligible if:

- They understood and spoke English; AND
- They had not attended *either* Triple P or Incredible Years parenting programmes prior to January 2018; AND
- They agreed that they and their child could participate in the research and that they would 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) was 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 could not be excluded prior to initial contact with practitioner, as the practitioner may be unaware of their psychological state. However, it was suggested that their inclusion was avoided if this information later comes to light as these parents may have difficulties engaging in the programme.

[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.

Both the delivery and evaluation team agreed, however, that it was 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

Our initial assumption (i.e. at proposal stage) was that the 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 believed it would be important to power to this level, even though this was 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 proposed 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.

MDES at proposal stage

		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
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		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

Due to recruitment challenges there were lower number of recruited settings and participants. The number of settings in the study in Summer 2019 was expected to be between 95 and 100. Furthermore, the number of participants per setting was expected to range between four and twelve. Under more conservative assumption, with 96 settings and eight participants per setting, the MDES was estimated to range between 0.254 and 0.266 for ICC=0.11 and ICC=0.13, respectively. The risk is that the trial was 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.

Update: In the end 67 settings were randomised. Taking into account the assumptions outlined above the MDES was estimated to range between 0.306 and 0.320 for ICC=0.11 and ICC=0.13, respectively (see table below). Given the risk that the trial would be underpowered, a second year of the trial had been planned, with the aim of recruiting a total sample of 150 settings in total across both trials. Due to Covid19 a second year of the trial was foregone.

MDES at randomisation

		ICC=0.11	ICC=0.13	PP
Minimum Detectable Effect Size (MDES)		0.306	0.320	0.469
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	34	34	34
	Control	33	33	33
	Total	67	67	67
Number of pupils	Intervention	272	272	68
	Control	264	264	66
	Total	536	563	134

Outcome measures

PRIMARY OUTCOME

The planned primary outcome for this evaluation was the child's expressive language. The rationale for having a particular focus on expressive language was 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., 2014; Rogde et al., 2016).

To measure expressive language, the plan was to use the Clinical Evaluation of Language Fundamentals for Preschool, Second Edition (CELF-P2). The general expressive language ability was to 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 Vocabulary [14] (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 score[15]. However, we had also planned to 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 was to be administered at the end of the trial in the week commencing 22 June 2020. Post-intervention outcome data was to be collected by Elklan[16] testers working for the evaluation team, who would not have knowledge of the allocation for each setting. This approach allows for blinding to allocation: we would 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 was to 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 was to be completed in three weeks, including one week of mop-up testing.

Update: Due to closure of early years settings as a result of national lockdown it was agreed that testing in settings was not feasible in Summer 2020. Follow-up with settings confirmed that testing in Autumn 2020 would not be feasible either as there was likely to be high measurement attrition given the fact many children had moved settings. Given the probability of high attrition and the minimal power associated with the given sample size (i.e. at randomisation) the decision was taken to not collect impact outcomes (i.e. CELF and SDQ-T).

SECONDARY OUTCOME

[14] The Sentence Structure test is based on 22 item and 0 to 1 scale; the Word Structure consist 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] Elkan 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/

The secondary outcome was to 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 agreed with Triple P UK, we planned to 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 a 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).

The plan was to collect both baseline and post-trial SDQ-T. for the children selected into the study before randomisation (based on the criteria above)[17]. Baseline testing was completed before randomisation, while post-trial SDQ –T was to be collected in June 2020. Collection of baseline data was mandatory for selected settings and families to be considered part of the trial. At baseline, the SDQ-T was administered by practitioners within the chosen early years settings. Baseline data collection was conducted by early years setting practitioners to reduce evaluation costs. As this occurred 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 decided to set up the SDQ on more user-friendly survey platform[18]. Additionally, if practitioners found it easier to read/respond to items on paper, we provided them with paper versions. Practitioners could therefore fill in either a paper or online version of the SDQ questionnaire. Although the scoring is simple, it could 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 **Appendix B**.

Update: In the summer term 2020, both control and intervention settings were to be asked to repeat the SDQ-T for those children that were originally selected. The staff were to receive online surveys. However, several reasons made testing with the SDQ-T unsatisfactory, including: low child attendance after settings re-opened in June (high attrition) and not enough time for staff to observe/interact with children before completing the questionnaire (unreliable data).

Analysis plan

The following analyses were planned but are no longer possible as no primary or secondary outcomes were collected post intervention: primary outcome analysis, secondary outcome analysis, subgroup analysis, effect size calculation, missing data analysis, sensitivity analysis. Details for the analyses that had been planned can be found in the original protocol[19].

IMBALANCE AT BASELINE

[17] 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
[18] We will use the Smartsurvey's platform.
[19] https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/Triple_P_protocol_final.pdf

We feel it will be beneficial to report on imbalance at baseline, despite the fact that the primary and secondary analyses have been forgone. 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

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

- Gender
- Average age in months
- SDQ T baseline score

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?

Update: Given the unknown impact of setting and school closures on families and settings, an additional research question was added to compliment the evaluation's understanding of results from the IPE:

RQ5: What is the impact of Covid-19 on Triple P eligible families and settings?

- How did settings support families over the period of setting closures?
- How did setting closures impact the delivery of Triple P?

Table 3 presents an overview of the IPE methods, including those that were updated.

Table 3: IPE methods overview

Research Methods	Data collection methods	Participants/data sources (type, number)	Data analysis methods	Research Questions addressed	Implementation / logic model relevance
Document analysis	Triple P Practitioner training materials (i.e. those provided as part of training in Triple P)	All Triple P training materials, e.g. power point, manual	Thematic analysis	RQ3, RQ4	Quality
Document analysis	Practitioner documentation prior to randomisation	All settings	Thematic analysis	RQ2	Fidelity, take-up, quality,
Surveys (4 time points)	Practitioner questionnaires, pre- and post-training, pre- and post-accreditation	Practitioner questionnaires collected by Triple P as part of programme	Thematic analysis, descriptive statistics	RQ2, RQ3	Quality, practitioner responsiveness
Document analysis	Parent attendance logs	Attendance logs from all Triple P sessions across all clusters	descriptive statistics, frequency counts	RQ2, RQ3	Compliance, take-up
Parent-report questionnaires	Parenting scale, DASS 21 (adult Depression, Anxiety and Stress Scale 21) and parents SDQ[20] administered by practitioners in	Aggregate responses by cluster of parent measures collected by Triple P	descriptive statistics, frequency counts	RQ2, RQ3	Reach and responsiveness

[20] 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

	treatment arm, pre and post				
Surveys	Online questionnaires	All practitioners, treatment settings	Descriptive statistics, frequency counts	RQ1, RQ2, RQ3	Fidelity, reach and responsiveness, quality, collect cost data
Surveys	Online questionnaires	All practitioners, control settings	Descriptive statistics, frequency counts	RQ4	Business as usual
Interviews	Semi structured telephone interviews	Practitioners in treatment settings (15)	Thematic analysis, deductive coding	RQ1, RQ2, RQ3, RQ5	Reach and responsiveness, impact of Covid-19 on settings and families, and delivery of Triple P
Interviews	Semi structured telephone interviews	Practitioners in control settings (10)	Thematic analysis, deductive coding	RQ4, RQ5	Business as usual, impact of Covid-19 on settings and families
Interviews	Semi structured telephone interviews	Parents in treatment settings (30)	Thematic analysis, deductive coding	RQ1, RQ2, RQ3, RQ5	Reach and responsiveness, quality, take-up, impact of Covid-19 on settings and families
Interviews	Semi structured telephone interviews	Parents in control settings (15)	Thematic analysis, deductive coding	RQ4, RQ5	Business as usual, impact of Covid-19 on settings and families

Key changes from the original IPE design are:

- No longer collecting control parent views of behaviour via text messages. Instead we conducted phone interviews with parents in the control arm (N=15) to understand their perceptions of their children's behaviour and to understand the impact of lockdown on their family
- Increased the number of interviews with practitioners in treatment arm to better understand their perceptions of the impact of Triple P on child and family outcomes (N=20)
- Increased the number of interviews with parents in the treatment arm (N=30)

1. Pre intervention phase (review of manual, reason for selecting children/families into the trial)

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

Selecting the appropriate children was a key part of the Theory of Change in terms of leading to better outcomes. Before randomisation practitioners were required to document the reason(s) why they had 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 SDQ[21] (child behaviour)

Practitioner questionnaires were administered at 4 different time points: pre- and post-training, pre- and post-accreditation. Triple P UK shared data with RAND Europe in aggregated form as this is a cost-effective approach that allowed us to use data from all the settings. The practitioners' confidence rating were aggregated at training cluster level (e.g. the regional areas identified in Table 2). Quantitative data from the practitioner questionnaires (e.g. practitioner confidence) will be summarised with results used to provide a description of practitioner responsiveness. This will give an insight at how effective training is at equipping practitioners to deliver the programme (i.e. quality).

Triple P UK aggregated the parent measures at training cluster level and shared these with RAND Europe. Quantitative data from the parent questionnaires (e.g. aggregate results on various measures at pre and post test) will be summarised with results used to provide a description of parent responsiveness. We will analyse results by training cluster but also compare them across training clusters to understand if key implementation factors (e.g. quality) are present.

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 were documented by practitioners delivering Triple P. The attendance logs were designed by the Evaluation team and they were easy to use. Attendance data from the parent attendance logs will be summarised with results used to provide a description of take-up and compliance of parents. This will give an insight at the extent to which the critical ingredients of the intervention are delivered to and/ or received by the target participants.

[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

Update: attendance at sessions was to be used in defining the compliance measure. However, as no evaluation-specific outcome measures were collected across treatment and control (i.e. as opposed to the intervention-related pre and post outcome measures collected as part of implementation by Triple P) this aspect of the evaluation has been forgone.

4. PRACTITIONER SURVEYS

We conducted online surveys with all practitioners in both arms of the trial at the end of the academic year. These were designed to collect perspectives of how the intervention had been implemented (in the treatment arm) and to understand what activities were conducted as business-as-usual (in the control arm)[22]. We included 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 prioritised understanding the activities of control settings as there was good implementation data collected by Triple P.

Descriptive findings from the survey in treatment settings will be aggregated and summarised in the report, with results used to provide a description of practitioner responsiveness, views on quality (i.e. of training and materials), fidelity to programme (RQ1, RQ2, RQ3). Additionally, any open-text responses were 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 was used to empirically examine the assumptions underlying the key mechanisms and processes 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). This will give an insight at how effective Triple P is at equipping practitioners to deliver the programme (i.e. quality), the ability of practitioners to deliver Triple P with fidelity, as well as identifying key barriers, facilitators and conditions needed to make Triple P succeed

Descriptive findings from the survey in control settings will be aggregated and summarised in the report, with results used to provide a description of business as usual in control settings. The approach will be similar to that used to analyse surveys of practitioners in treatment settings outlined above. However, the focus will be on providing a description of the counterfactual based on the responses from practitioner surveys in control settings (RQ4).

[22] 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.

Practitioners were 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).[23] Questions were designed to ensure adherence to the EEF's guidance and collected data on both financial costs and practitioner time.

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

5. PHONE INTERVIEWS WITH PARENTS

We conducted 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 proposed using a stratified sampling (see McDavid et al., 2019) to draw views from a meaningful sample of parents in both the treatment and control settings. Once the settings had been chosen, practitioners were asked to identify parents that we could approach for interview. More information on the sampling approach for identifying settings is described in more detail below. Practitioners were 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). We offered 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 Husain, et al. 2018).

Parent interviews with parents in treatment settings focused on examining the intervention mechanisms of change, particularly any changes to the home learning environment and child language development in relation to parenting strategies as well as the impact of Covid-19 on their family. In designing an interview protocol, we have consulted 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 also examined parent preferences for programme delivery (Sonuga-Barke et al., 2018).

Parent interviews with parents in control settings focused on examining the behaviour of parents to better describe business as usual and provide a comparison to behaviours of parents in treatment settings. Include in the interview protocol were questions aimed at understanding the home learning environment and child language development in relation to parenting strategies as well as the impact of Covid-19 on their family.

In analysing interviews in both treatment and control settings a general inductive approach was 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.

Update: we included interviews with parents in the control arm. This was to replace the texts to parents and was seen as a richer means of collecting data on the impact of Covid-19 on families, as well as a means to establish a very light-touch counterfactual to compare to parents in the treatment arm.

[23] 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.

6. **UPDATE:** PHONE INTERVIEWS WITH PRACTITIONERS IN TREATMENT AND CONTROL ARMS

We conducted phone interviews with practitioners from both arms, as phone interviews have the benefit of being cost effective and minimise time burden on practitioners.

We used a stratified sampling (see McDavid et al., 2019) to draw views from a meaningful sample of practitioners across both arms. The aim was to establish equal numbers of interviewees from each cluster. We randomly selected settings from each training location (cluster 1 through cluster 6) in order to allow settings from different training locations to be studied in more depth.

Update: we intended to first rank the settings according to frequency of parent attendance (i.e. high, medium, low), using the parent attendance logs. However, delays in collecting parent attendance logs from settings meant that this aspect of the evaluation has been forgone to ensure that practitioners could be interviewed in a timely manner (e.g. before summer holidays).

Practitioner interviews focused on examining the intervention mechanisms of change, implementation factors, perceptions of Triple P, practitioner responsiveness as well as the impact of Covid-19 on their settings and families in their setting. In designing an interview protocol, we consulted 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. As with parents a general inductive approach was used for analysis.

Cost evaluation

Cost data was gathered through practitioner online surveys in the implementation and process evaluation. Questions were 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 used this information to estimate cost per-child, following EEF guidelines (EEF, 2015).

The main costs of the intervention related 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 relied on data provided by the delivery team. RAND Europe also considered the cost of the time of practitioners and other staff in delivering the programme.

Questions were 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 used the information on direct and indirect costs to estimate cost per-child, following EEF guidelines (EEF, 2019).

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.

Update: In May the privacy notice was updated to reflect the rescoped data collection activities (including foregoing outcome testing). This was shared with all participants and schools in both arms of the study.

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 it subcontract. Data obtained by Elklan will be obtained 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: 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 atypically long delay between parent recruitment and group delivery. Mitigation provision of retention support materials to intervention settings. The study is powered to detect and 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
4th October	Teacher SDQ, children and parent Information (baseline data) completed and shared with RAND	Setting practitioners
September-October (by October 9th)	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
23rd October	Implementation Planning Consultation begins with Intervention Sites.	Triple P UK /Settings
Mid October 2019	Practitioner Training resources arrive to venues	Triple P UK
4th- 8th 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 27th week	Clinical workshops	Triple P UK / Setting
February 3rd week (break for half term 10th April 2020	Group Triple P delivery Collection of parental attendance logs	Setting
March 2020	Closure of early years settings to most children as a result of national lockdown	
June-September 2020	Online surveys with practitioners and parents Interviews with parents	RAND Europe
July 31st 2020	Parent Delivery completed	Setting
September 2020	Parent pre- and post-data scored centrally / shared with RAND	Triple P UK / Setting
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.

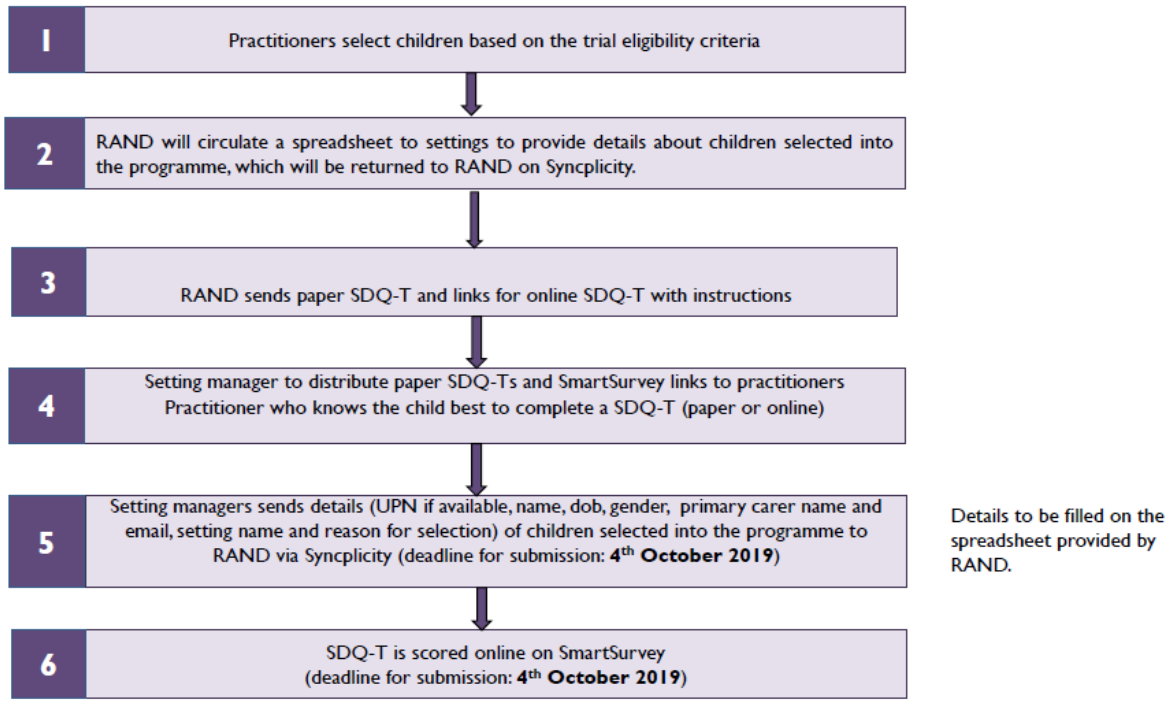
Session 5-7: The practitioner provides feedback from initial assessments that the family completed and then uses the self-regulatory feedback model to help parents review their implementation of planning ahead routines for their high-risk situations. From this, parents set goals for further refinement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

Allen, G. (2011). *Early intervention: the next steps, an independent report to Her Majesty's government by Graham Allen MP*. [pdf] London: The Stationery Office. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/284086/early-intervention-next-steps2.pdf [Accessed 25 September 2020].

Asmussen, K., & Weizel, K. (2010). *Evaluating the evidence: Fathers, families and children*. [pdf] London: National Academy for Parenting Practitioners. Available at: http://pupilpremium.learningtrust.net/uploads/1/2/7/4/12746193/fathers_families_and_children.pdf [Accessed 25 September 2020].

Bandura, A., 1977. Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review*, 84(2), pp. 191-215.

Bandura, A. (Ed.), 1995. *Self-efficacy in changing societies*. Cambridge: 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), pp. 932-942.

Campbell, M. K., Piaggio, G., Elbourne, D. R., & Altman, D. G., 2010. Consort 2010 statement: extension to cluster randomised trials. *British Medical Journal*, 345, p.e5661.

Cann, W., Rogers, H., & Worley, G. (2003) Report on a program evaluation of a telephone assisted parenting support service for families living in isolated rural areas. *Australian e-Journal for the Advancement of Mental Health*, 2(3).

Carril, A., 2017. Dealing with misfits in random treatment assignment. *Stata Journal*, 17(3), pp. 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), pp. 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), pp. 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), pp. 11-13.

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), pp. 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), pp. 1210-1219.

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), pp. 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), pp. 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), pp. 772-781.

Early Intervention Foundation, 2016. *Foundations for life: what works to support parent child interaction in the early years.* [pdf] London: Early Intervention Foundation. Available at https://www.eif.org.uk/files/pdf/foundations-for-life.pdf [Accessed 25 September 2020].

Education Endowment Foundation, 2015. EEF Guidance on Cost Evaluations. London: Education Endowment Foundation. [online] Available at: https://v1.educationendowmentfoundation.org.uk/uploads/pdf/EEF_guidance_to_evaluators_on_cost_evaluation.pdf

Education Endowment Foundation (2018) *Statistical Guidance for EEF evaluations*. [pdf] London: Education Endowment Foundation. Available at: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol_or_SAP/EEF_statistical_analysis_guidance_2018.pdf [Accessed 25 September 2020]

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), pp. 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), pp. 1-172.

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), pp. 90-110.

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

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

Gould, N. & Richardson, J., 2006. 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), pp. 47-60.

Gray, G. R., Totsika, V., & Lindsay, G., 2018. Sustained Effectiveness of Evidence-Based Parenting Programs After the Research Trial Ends. *Frontiers in Psychology*, 9, pp. 2035.

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

Higgins, J. P. T., Savović, J., & Page, M. J., 2016. *Revised Cochrane Risk of Bias Tool for Randomized Trials (RoB 2.0)*. [online] Available at: https://sites.google.com/site/riskofbiastool/welcome/rob-2-0-tool [Accessed 25 September 2020]

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), pp. 191-206.

Husain, F., Wishart, R., Marshall, L., Frankenberg, S., Bussard, L., Chidley, S., Hudson, R., Votjkova, M. & Morris, S., 2018. *Family Skills Evaluation Report*. London: 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.

Kim, J. H., Schulz, W., Zimmerman, T., & Hahlweg, K., 2018. Parent-child interactions and child outcomes: Evidence from Randomized Intervention. *Labour Economics*, 54(2018), pp. 152-171.

Leung, C., Sanders, M., Leung, S., Mak, R., & Lau, J., 2003. An Outcome Evaluation of the Implementation of the Triple P-Positive Parenting Program in Hong Kong. *Family Process*, 42(4), pp. 531-544.

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, pp. 11-24.

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), pp. 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), pp. 677-692.

McDavid, J. C., Huse, I. & Hawthorn, L.R.L., 2019. *Program Evaluation and Performance Measurement*. Thousand Oaks: Sage Publications.

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), pp. 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), pp. 114-144.

Oakes, J. M., 2013. Effect identification in comparative effectiveness research. *The Journal for Electronic Health Data and Methods*, 1(1), pp. 1-7.

Pascal, C., Bertram, T., Cullinane, C., & Holt-White, E. (2020). *Covid-19 and Social Mobility Impact Brief #4: Early Years*. [pdf] London: The Sutton Trust. Available at: https://www.suttontrust.com/wp-content/uploads/2020/06/Early-Years-Impact-Brief.pdf [Accessed 25 September 2020].

Patterson, G. R., Chamberlain, P., & Reid, J. B., 1982. A comparative evaluation of a parent-training program. *Behavior Therapy*, 13(5), pp. 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), pp. 427-440.

Portfield, J. K., Herbert-Jackson, E., & Risley, T. R., 1976. Contingent observation: an effective and acceptable procedure for reducing disruptive behavior of young children in a group setting. *Journal of Applied Behaviour Analysis*, 9(1), pp. 55-64.

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), pp. 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), pp. 544-557.

Robinson-Smith, L., Merrell, C., Stothard, S., Torgerson, C., 2018. *Independent Evaluation of the EasyPeasy programme*. [online] Available at: https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/easypeasy-learning-through-play/ [Accessed 25 September 2020].

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), pp. 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), pp. 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), pp. 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), pp. 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), pp. 506-517.

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), pp. 337-357.

Scott, S. (2010). National dissemination of effective parenting programmes to improve child outcomes. *The British Journal of Psychiatry*, 196(1), pp. 1-3.

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), pp. 816-827.

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), pp. 797-809.

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), pp. 605-627.

Thomas, D. R., 2006. A general inductive Approach for analyzing qualitative evaluation data. *American Journal of Evaluation*, 27(2), pp. 237 – 246.

Wiig, E. H., Secord, W., & Semel, E., 2006. *Clinical evaluation of language fundamentals – Preschool second UK edition*. Harcourt Assessment.

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), pp. 130-146.

World Health Organization, 2016. *INSPIRE Seven Strategies for Ending Violence Against Children*. [pdf] Geneva: World Health Organization. Available at: https://www.who.int/publications/i/item/inspire-seven-strategies-for-ending-violence-against-children [Accessed 25 September 2020].

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), pp. 287-304.