# Supporting Parents & Kids Through Lockdown Experiences **(SPARKLE)** Trial

A parallel randomised controlled trial of a digital parenting support app implemented in the general population during the COVID-19 pandemic

Statistical Analysis Plan
Version 1.1 started 12/10/2021

Refers to Protocol Version 1.1, filename: SPARKLE protocol V1.1 20.07.2021.docx

Clinicaltrials.gov: NCT04786080

Trial Statistician: Nicholas Beckley-Hoelscher

Signature.....  ..... Date...16/10/2021.....

Chief Investigator: Edmund Sonuga-Barke

Signature....  .Date 17.10.21

Senior Statistician: Kimberley Goldsmith

Signature.....  ..... ……..
Date.19.10.2021

Trial Steering Committee Chair: Professor Saskia Van der Oord

Signature...  Date...18/10/2021

## CONTENTS

This document contains up to date statistical analysis plans (with version numbers and dates), including a quantitative analysis plan and a schedule of assessments and measures.

This SAP refers solely to the main clinical effectiveness analysis of the primary and secondary trial outcomes, i.e. the analysis to produce the main trial results paper. This document does not contain details on the health economic or any other analysis plan. Specifically, it does not address objectives iv, v and vi on p5 of the protocol.

# QUANTITATIVE ANALYSIS PLAN

Katarzyna Kostyrka-Allchorne (Co-I). Department of Child & Adolescent Psychiatry, Institute of Psychiatry, Psychology & Neuroscience, King's College London (kasia.kostyrka-allchorne@kcl.ac.uk).

Cathy Creswell (Co-I). Departments of Psychiatry and Experimental Psychology, University of Oxford (cathy.creswell@psych.ox.ac.uk).

Sarah Byford (Co-I – Senior Trial Health Economist). Department of Health Service and Population Research, Institute of Psychiatry, Psychology & Neuroscience, King's College London (sarah.byford@kcl.ac.uk).

Crispin Day (Clinical collaborator). Department of Psychology, Institute of Psychiatry, Psychology & Neuroscience, King's College London (crispin.1.day@kcl.ac.uk).

Kimberley Goldsmith (Co-I – Senior Trial Statistician). Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology & Neuroscience, King's College London (kimberley.goldsmith@kcl.ac.uk).

Marta Koch (SPARKLE Research Administrator). Department of Child & Adolescent Psychiatry, Institute of Psychiatry, Psychology & Neuroscience, King's College London (marta.koch@kcl.ac.uk).

Walter Muruet Gutierrez (Junior Trial Statistician). Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology & Neuroscience, King's College London (walter.muruet_gutierrez@kcl.ac.uk).

Melanie Palmer (SPARKLE Research Fellow). Department of Child & Adolescent Psychiatry, Institute of Psychiatry, Psychology & Neuroscience, King's College London (melanie.palmer@kcl.ac.uk).

Olly Robertson (SPARKLE Research Assistant). Departments of Psychiatry and Experimental Psychology, University of Oxford (olly.robertson@psych.ox.ac.uk).

James Shearer (Junior Trial Health Economist). Department of Health Service and Population Research, Institute of Psychiatry, Psychology & Neuroscience, King's College London (james.shearer@kcl.ac.uk).

Petr Slovak (HCI collaborator). Department of Child & Adolescent Psychiatry, Institute of Psychiatry, Psychology & Neuroscience, King's College London (petr.slovac@kcl.ac.uk).

Polly Waite (Co-I). Departments of Psychiatry and Experimental Psychology, University of Oxford (polly.waite@psych.ox.ac.uk).

Edmund J S Sonuga-Barke (PI - corresponding author). Department of Child & Adolescent Psychiatry, Institute of Psychiatry, Psychology & Neuroscience, King's College London (edmund.sonuga-barke@kcl.ac.uk).

## AMENDMENTS TO THE SAP

| Date | SAP version | Protocol version and date | Reason for change/Amendment made |
|---|---|---|---|
| 16/08/2021 | V1.0 | Version 1.1, filename: SPARKLE protocol V1.1 10.07.2021.docx | Initial version |
| 23/09/2021 | V1.01 | Version 1.1, filename: SPARKLE protocol V1.1 20.07.2021.docx | • More information about blinding included<br>• Additional descriptive statistics for variables with many zero values<br>• Description of one-sided confidence intervals<br>• Updated sections on model checks and sensitivity analyses |
| 12/10/2021 | V1.1 | Version 1.1, filename: SPARKLE protocol V1.1 20.07.2021.docx | Added reporting of percentage of participants with non-zero app usage data (section 2.7) |

# 1. Description of the trial

The COVID-19 related lockdowns and continuing social distancing measures have presented families with unprecedented challenges. Data from a nation-wide cohort study tracking changes in families' mental health during the lockdown (Co-SPACE) show a significant rise in parent-reported children's behaviour and associated family-related stress (Waite, Patalay et al. 2020). Furthermore, there is also evidence that the majority of parents participating in Co-SPACE want additional support (Waite, Patalay et al. 2020). SPARKLE is a two-arm superiority parallel-group randomised controlled trial with some characteristics consistent with a trial within a cohort (TwiC, Relton et al 2010) that will examine whether the use of a digital public health intervention (Parent Positive) can reverse the unfavourable effects associated with the COVID-19 pandemic on children's conduct problems and family conflict.

Parent Positive is a mobile phone application for the most commonly used smartphone platforms (iOS and Android) to support parents in managing their children's behaviour. The application is based on extensive parenting research and was co-designed with parents. It provides a flexible digital space where parents can access advice, a peer-support platform and a curated list of evidence-based parenting resources. It consists of three zones:

i) The *Parenting Boosters* zone includes structured advice, support and tips for parents to deal with eight common parenting challenges. These are based around the series of Families Under Pressure animations identified and developed through co-design with parents. To increase their attractiveness and reach these were scripted to be light-hearted, humorous and non-judgmental and are delivered by eight high-profile British celebrities who are also parents. The eight messages relate to: (i) staying positive and motivated (Olivia Colman); (ii) making sure everyone knows what is expected of them (Sharon Horgan); (iii) building your child's self-confidence

and trust (Danny Dyer); (iv) getting your child to follow instructions (Rob Brydon); (v) promoting better behaviour (Jessica Ennis-Hill); (vi) limiting conflict (Holly Willoughby); (vii) keeping calm when your kids act up (Romesh Ranganathan); and (viii) careful use of sanctions (Shappi Khorsandi).

ii) The *Parenting Exchange* zone provides a facilitated parent-to-parent communication platform where parents are encouraged to raise any specific challenges they are facing to receive support from other parents. Trained parent facilitators will moderate the exchange and create posts to enhance engagement with the intervention. The Exchange will be also used to collate questions that parents have for experts (e.g., developmental psychologists, parent training practitioners, nutritionists, etc.) about a range of topics related to both the *Parent Positive* challenges and broader issues of direct relevance to children's behaviour (e.g., sleep, diet, etc). These questions will be answered by experts at pre-recorded webinars held regularly during the period of the trial. The recording will be made available to parents using the app.

iii) The *Parent Resources* zone will provide links to carefully selected high-quality, evidence-based online parenting resources reviewed and approved by a committee of parenting experts.

The app will be free and parents will be able to access the information when needed and in the order they choose. Parents in the intervention group will receive access during the immediate post-randomisation period until 30th November 2021. To access the app, parents will receive an automated email with a link to download it from either Google Play (for Android users) or the App Store (for Apple users) together with brief instructions on how to download the app on the smartphone and register as a user.

SPARKLE is embedded within Co-SPACE (Waite and Creswell 2020). The trial will recruit parents with children between the ages of 4 and 10 interested in participating and who have a device compatible with the Parent Positive app. Participant allocation to intervention (Parent Positive) or control (Follow-up as usual, FAU) groups will be performed using simple randomisation in a 1:1 ratio. Further details are available in the study protocol V1.0 called SPARKLE protocol V1.0 15.03.2021.docx, available for download here.

## *1.1 Principal research objectives to be addressed*

Primary objective

To evaluate Parent Positive's effectiveness to reduce the levels of parent-reported child conduct problems compared to follow-up as usual at one-month post-randomisation.

Secondary objectives

To investigate the effectiveness of Parent Positive compared to follow-up as usual on:

1. Reducing the levels of parent-reported child conduct problems at two months post-randomisation follow-up.
2. Reducing levels of parent psychological distress at one and two months post-randomisation follow-up.

3. Reducing levels of parental child-related stress and worries at one and two months post-randomisation follow-up.
4. Reducing levels of family conflict at one and two months post-randomisation follow-up.
5. Reducing levels of child emotional problems at one and two months post-randomisation follow-up.
6. To assess the extent to which the effects of Parent Positive on the primary conduct outcome are moderated by levels of (i) pre-existing conduct problems and (ii) *Parent Positive* usage as monitored during the one month intervention period.
7. To evaluate the cost-effectiveness of Parent Positive at two-month post-randomisation follow-up. A separate Health Economic Analysis plan will address this objective; it will not be addressed in this SAP.

Exploratory Objectives

We intend to investigate whether the effects of Parent Positive's use on the primary and secondary outcomes are moderated by other baseline characteristics - family socioeconomic status and composition, parental psychological distress, child age, symptoms of attention-deficit/hyperactivity disorder and emotional problems and by baseline lockdown circumstances and policies at baseline on the effects of *Parent Positive* on primary and secondary outcomes. Data on lockdown circumstances will be collected at one and two months post-randomisation for descriptive purposes. Note that these exploratory aims will not be addressed in the primary trial results paper and so the associated analysis is not covered in this SAP.

## *1.2 Trial design including blinding*

This study is a two-month, parallel-arm, superiority randomised controlled clinical trial comparing the effect of Parent Positive as compared to FAU, with a primary outcome of parent-reported child conduct problems at one month post-randomisation. The study will be embedded into an existing UK-wide, large, self-selected community cohort (Co-SPACE), with all current and new participants aged ≥ 18 years who have children aged 4-10 years invited to take part in the trial. Those parents/carers who consent to SPARKLE will be assigned by simple 1:1 randomisation to Parent Positive or FAU groups. Randomisation will be carried out remotely within Co-SPACE in the Qualtrics platform – a web-based survey and data collection software platform – using the "Randomizer" function and followed up using a separate survey branch. Qualtrics will automatically inform the parent about their group allocation. A researcher responsible for providing participants with instructions regarding the app access will obtain information about the outcome of randomisation from the Qualtrics database. Outcome data will be collected remotely via Qualtrics. Outcome measures will be collected according to the Co-SPACE schedule at baseline, which will be the Co-SPACE survey data obtained immediately prior to randomisation, and then at one month and two months post-randomisation. After the second post-randomisation time point, parents will continue to be involved in Co-SPACE according to the Co-SPACE data collection schedule unless consent is withdrawn.

OR and AS will have access to the Qualtrics system through which randomisation occurs so it will not be possible for them to be blind. Bias associated with unblinding is low, given the online nature of data collection for the trial.

MP and MK will have access to Parent Positive usage data so will not be able to be blind. OR will also have access to the Parent Positive usage data. Bias associated with unblinding is low, given the online nature of data collection for the trial.

The following members of the research team will be blinded through until after the analysis is complete unless unblinding is required for some reason, such as if the TSC expressed concern over adverse events: ES-B, KK-A, CC, SB, CD, PS, and PW. If the TSC expresses concern over adverse events, unblinding will occur for the minimum number of the research team possible.

NB-H, the junior statistician, and JS, the junior health economist, will be unblinded throughout. KG, the senior statistician, will remain blind until the initial draft of the statistical report has been produced.

## *1.3 Eligibility screening*

Eligible individuals for, or current participants of, Co-SPACE will be invited to participate in SPARKLE. The inclusion criteria for Co-SPACE are: The parent is willing and able to give informed consent, must be at least 18 years old and living in the UK. There are no additional exclusion criteria.

### 1.3.1 Inclusion criteria

1. Parents aged ≥ 18 years
2. Has a child aged 4 to 10 years
3. Access to a smartphone with operating system OS 8-9 or higher (Android devices) or iOS 12-13 or higher (Apple devices)

### 1.3.2 Exclusion criteria

1. No further exclusion criteria

## *1.4 Method of allocation of groups*

Once baseline assessments are complete, the individuals will be randomised 1:1 to the Parent Positive and FAU groups. Randomisation will be at patient level, using the Randomiser tool of the online Qualtrics platform and followed up using a separate survey branch from those in the Co-SPACE study. Simple randomisation will be used; neither blocking nor stratification will be employed. Randomisation will occur automatically after receiving informed consent, and participants will receive an immediate notification in the survey platform and via email of their group allocation.

A researcher responsible for providing participants with instructions regarding the app access will obtain information about the outcome of randomisation from the Qualtrics database.

## *1.5 Duration of the treatment period*

Participants allocated to Parent Positive will have access to the app until 30th November, 2021. Those individuals randomised to the FAU will be given access to the app after about two-months post-randomisation and after baseline and the two post-randomisation time points are collected, and then will also be able to access Parent Positive until the 30th of November 2021.

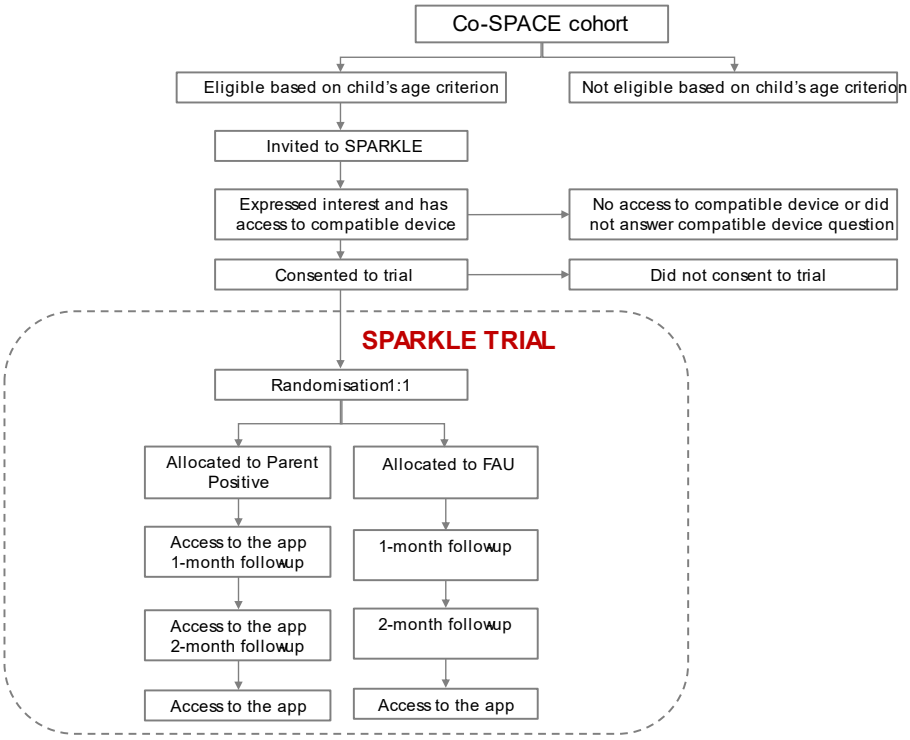
## *1.6 Frequency and duration of follow-up*

The enrolment and group allocation procedures are shown in Figure 1. Baseline measures will correspond to the data collected by Co-SPACE immediately before randomisation. Assessment of follow-up measures will occur at one and two months post-randomisation.

Please see SCHEDULE OF ASSESSMENTS AND MEASURES for details and the data collection schedule.

### *Figure 1. Trial design flow diagram*

Schematic diagram of flow of participants through the trial. Please see the CONSORT diagram (Figure 2) for detail on how this will be reported.




## *1.7 Survey completion timepoints*

An overview of the trial schedule is shown in Figure 1. Trial design flow diagram.

Baseline measures will consist of data routinely collected by Co-SPACE immediately before randomisation. Participants will be able to access the baseline questionnaire an unlimited number of times up to 7 days (inclusive) afterwards. Participants will be sent email and text reminders to do their one- and two-month post-randomisation follow-up questionnaires remotely via Qualtrics. Questionnaires will become available/reminders will be sent on the appropriate day post randomisation (one month for the first follow-up, two months for the second). Participants will be able to access the follow-up questionnaires an unlimited number of times from the moment they become available and up to 14 days (inclusive) afterwards. Once surveys are submitted, answers cannot be changed.

Please see Table 1 under SCHEDULE OF ASSESSMENTS AND MEASURES for details and the data collection schedule.

## *1.8 Measures*

### 1.8.1 Baseline measures

The following measures are recorded at baseline: Demographic features and family characteristics; Strengths and Difficulties Questionnaire (SDQ)(Goodman 1997); parental child-related stress and worries and family conflict, which are both scales designed specifically for the Co-SPACE study; Depression, Anxiety, and Stress Scale - 21 items (DASS-21 converted to DASS-42 equivalent scoring of the DASS-21

items) (Antony, Bieling et al. 1998); and lockdown circumstances. Please see Appendix 1 – Scoring of outcome measures for scoring detail. Note that the full SDQ will be gathered for the purposes of the health economic analysis, but only the conduct problems and emotional problems subscales are clinical effectiveness outcomes dealt with in this SAP. The three other SDQ subscales (hyperactivity-inattention, peer problems, prosocial) will be summarised at baseline for the primary paper (either by the health economists or statisticians).

Demographic and family characteristics recorded at baseline are: child's age, gender, parent and child's ethnicity, number of adults in the household, number of children in the household, number of people living in the household, number of rooms in the family home (excluding bathrooms or toilets), and access to outside space. A variable indicating single adult household yes/no is derived. An overcrowding index is calculated as the total number of people in the household divided by the number of rooms. A variable indicating an overcrowded household yes/no is derived from this index as >1 is overcrowded, <=1 is not overcrowded. Please see Appendix 1 – Scoring of outcome measures for more detail on the coding of the household and overcrowding variables. We will estimate family SES based on total household income, coded as < £16,000 | £16,000 - £29,999 | £30,000 - £59,999 | £60,000 - £89,999 | £90,000 - £119,999 | > £120,000 | Prefer not to say, likely collapsing small categories. Lockdown circumstances to be recorded include whether the parent, child, household members, close friends or family outside the household have had COVID-19 in the last month, current isolation status, whether their area has been in local/national lockdown, whether parents are working from home, and whether or not the children have been attending school. Note, the Co-SPACE study has been using government dates to designate full/partial/no lockdown rather than the self-report of whether the area has been in lockdown variable, SPARKLE will do the same.

### 1.8.2 Primary outcome measure

The primary outcome measure is the parent-reported level of child conduct problems at one-month post-randomisation. Conduct problems will be measured using the conduct problems sub-scale of the SDQ. Please see Appendix 1 – Scoring of outcome measures for scoring detail.

### 1.8.3 Secondary outcome measures

Secondary outcomes are:

1. Parent reported level of child conduct problems at two months post-randomisation as measured by the conduct sub-scale of the SDQ.
2. Parent reported level of child emotional problems at one and two months post-randomisation as measured by the emotional symptoms sub-scale of SDQ.
3. Parent psychological distress at one and two months post-randomisation as measured by the DASS-42 equivalent scoring of the DASS-21 items (see Appendix 1 – Scoring of outcome measures). We will refer to this throughout the rest of the main SAP document as DASS.
4. Parental child-related stress and worries levels at one and two-months post-randomisation as measured by a five item sub-scale in Co-SPACE.
5. Levels of family conflict at one and two months post-randomisation as measured by a three item sub-scale used in Co-SPACE.

### 1.8.4 Mediators of treatment effects

No mediation analysis is planned.

### 1.8.5 Moderators of treatment effects

To address secondary objective 6 (Section 1.1), the effects of the baseline level of child conduct problems will be explored as a moderator, and the usage of Parent Positive between randomisation and one month post-randomisation (please see Section 1.8.6 for the specific measure) will be explored as a post-randomisation effect modifier of the Parent Positive vs FAU treatment effect on the primary outcome, i.e. child conduct problems at one-month post-randomisation. Although not specified in the protocol, it is of interest to study these effects on the secondary outcome of child conduct problems at two-months post-randomisation - we will do this and add to any future protocol amendment.

### 1.8.6 Other measures

Attitudes to Parent Positive and measures of usage of the Parent Positive app will also be collected, in the intervention group only, with usage values of zero entered into the database for participants in the control arm. Attitudes will be measured using the Parent Positive Evaluation Questions, please see Appendix 1A for a copy of this measure.

Specifically, attitudes will be measured using the three items asking about parents' ratings of the usefulness of each of the Booster, Exchange and Resources zones rated on a 7-point scale (0 = not useful at all to 6 = very useful), which will be averaged to derive an overall app usefulness score. See Appendix 1A, Parent Positive Evaluation Questionnaire, these are the questions called, Booster_Rating, Exchange_Rating, Resources_Rating. The questionnaire will also capture parents' descriptive and qualitative views of Parent Positive. The statisticians will only summarise the three usefulness responses and the overall usefulness score, but not any of the other responses, which will be dealt with by other members of the study team and may or may not be presented in the primary paper.

Usage data will be available for the Parent Positive Booster and Exchange zones. There are no quantitative usage variables associated with the Resources zone, so it will not be dealt with further in this SAP. Usage will be summarised separately for the randomisation (T1) to one month post-randomisation (T2) period, and the one month post-randomisation (T2) to two months post-randomisation (T3) period. We note that the app usage data will be gathered cumulatively at T2 and T3, so variables for T2 to T3 will need to be derived by subtracting the usage up to T2 from usage at T3. The app usage variables of interest are:

1. How many engaged with the app at all,
2. How many times was the app accessed,
3. How many viewed at least one of the boosters,
4. How many boosters were entered,
5. How much time was spent in each of the eight boosters individually,
6. How much time was spent across all eight boosters (sum of the time spent in each individual booster), and did this vary by month of randomisation into the trial. Rather than using calendar month, we will use one month time periods starting from 19/05/2021 (the first day of randomisation),
7. How many published a post (one or more) in the Exchange zone, and how many posts were made,

8. How many commented on a post (one or more) in the Exchange zone, and how many comments were made,
9. How many expert videos were watched in the Exchange zone. **NB: at the time of SAP writing it is known that this variable has not been collected correctly in the app for a large proportion of the participants, so this variable may not be used/summarised**,

with #6 up to T2 being the variable used to assess post-randomisation effect modification by app usage (individuals not using the app/boosters will be coded as having spent zero time across all boosters; please see Sections 1.8.5 and 3.1.3.2).

### *1.9 Sample size estimation*

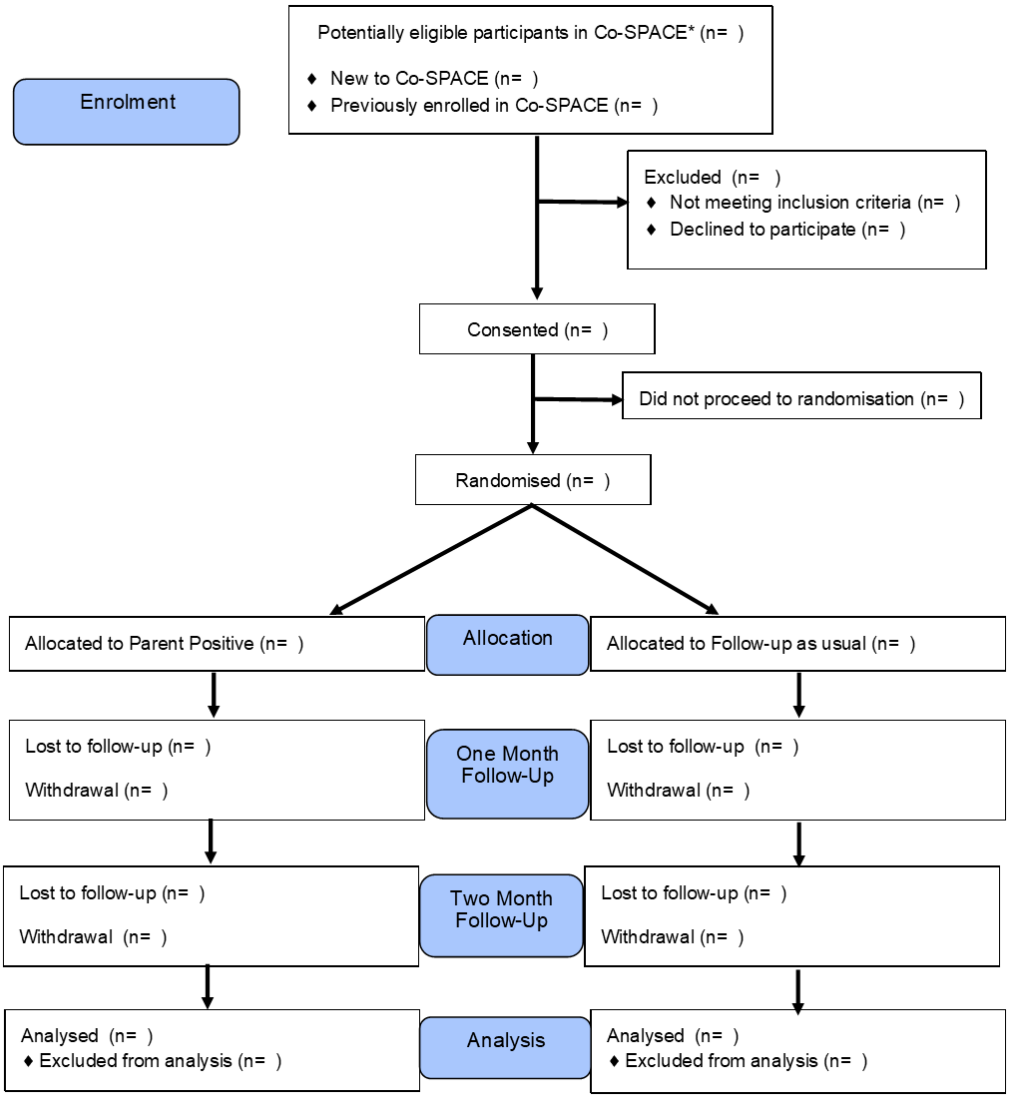
A total of 616 will be recruited into the trial with 308 consenting parents randomised to each group following SPARKLE baseline assessments. This sample size was powered to address our primary question whether comparing outcomes at one month post-randomisation, *Parent Positive* will reduce children's conduct problems observed as compared to FAU. During the first UK lockdown (i.e., March to June 2020) there was around an approximate 0.2 standard deviation increase in SDQ conduct problems scores reported by Co-SPACE parents with children aged 4-10. For the current study, we took this as representing the smallest between group difference that was of clinical value to detect between *Parent Positive* and FAU scores post-randomisation (i.e., Cohen's *d* of 0.2). We assumed a within trial drop-out rate of 30%, a correlation of 0.5 between one pre- and two post-randomisation measures (Machin, Campbell et al. 2018), and using a one-tailed (Parent Positive > FAU) and an alpha of .05 that this number of participants provides 90% power to test the one-sided hypothesis that *Parent Positive* is superior to FAU.

## 2. Data analysis plan – Data description

### *2.1 Recruitment and representativeness of recruited patients*

A CONSORT diagram (see Figure 2 for a schematic example, the final version may differ) will be constructed (Moher, Schulz et al. 2001). Parents are asked on entering CO-SPACE to confirm they are over the age of 18, so the initial number of potentially eligible participants will be Co-SPACE study parents with children of eligible age for SPARKLE. We will then report the number of parents further eligible based on having a compatible device and consenting to SPARKLE, then, by intervention group, the number of participants continuing through the trial, the number withdrawing, the number lost to follow-up and the number of excluded/analysed.

***Figure 2. Template CONSORT diagram for SPARKLE trial***




*eligible based only on child's age only at this stage, not on whether participant has a compatible device

## *2.2 Baseline comparability of randomised groups*

The description of variables measured at baseline is in section 1.8.1. Baseline variables will be described by intervention group and overall. Frequencies and proportions will describe categorical variables, while numerical variables will be described using mean and standard deviation (SD) if normally distributed and median and interquartile range (IQR) if not normally distributed. Baseline variables to be described will be child's age, gender and ethnicity, parent's ethnicity, number of adults in the family, number of children in the family, total number of people in the household, single parent household yes/no, number of rooms in the family home, overcrowding index, overcrowding yes/no, access to outside space yes/no, socioeconomic status as the categorical income variable, the lockdown status variables as described in section 1.8.1 with whether in lockdown or not derived from government dates and policies rather than using the self-report variables, parent reported SDQ conduct problems, child emotional problems, hyperactivity-inattention, peer problems and prosocial sub-scale scores (the latter three to provide baseline information on the sample), parental child-related stress and worries score, family conflict score, and the parental DASS score.

No statistical testing of the baseline differences between randomised groups will be done.

## *2.3 Treatment adherence and withdrawal from treatment and trial*

Because the sample is not restricted to those with a specific clinical need, participants do not have to adhere to a specific treatment protocol and there is not a definition of treatment adherence as such, nor a definition of withdrawal from treatment. However, an analysis of the extent to which effects depend on intervention uptake will be conducted (please see sections 1.8.5 and 3.1.3.2). We will also summarise the other measures in section 1.8.6 as follows.

The individual app usefulness measures and averaged overall score will be summarized using the mean and SD and median and IQR.

The app usage measures will be summarised as follows. For #2, 4-6, these will be summarised both among those using the app and for everyone in the Parent Positive group (i.e. the latter will code zero values for those in the Parent Positive group who do not use the app). For #7 and 8, these will be summarised only among those using the app.

1. Number using the app at all, and the proportion out of the total number given access to app in the Parent Positive group (taken from the number of times the app was accessed, with one time accessing or more equating to using the app at all),
2. Mean (SD), median (IQR) and range of number of times the app was accessed,
3. Number viewing any boosters, and the proportion out of the number using the app (in #1),
4. Mean (SD), median (IQR) and range of number of boosters started (taken from the number of boosters individuals spent time in),
5. Mean (SD), median (IQR) and range of time spent in each of the eight boosters individually,

6. Mean (SD), median (IQR) and range of time spent across all eight boosters (sum of the time spent in each individual booster), overall and by one month time periods starting from 19/05/2021 (the first day of randomisation to see if this looks to vary over time (i.e., if the app is used less at the beginning of the study while issues with the app are being ironed out, the app is used less or more in school holiday months, etc.) ,
7. Number publishing >= 1 post in the Exchange zone, and the proportion out of those engaging with the app in #1; and mean (SD), median (IQR) and range of number of published posts,
8. Number posting >= 1 comment on a post in the Exchange zone, and the proportion out of those engaging with the app in #1; and mean (SD), median (IQR) and range of number of comments.

The number and proportion withdrawing from the trial (i.e., actively state they are unwilling to provide any further research data), and the reasons for withdrawal will be summarised by intervention group and overall.

### *2.4 Loss to follow-up and other missing data*

The number and proportion of participants missing each primary and secondary outcome variable will be summarised by intervention group and overall at each time point.

### *2.5 Adverse event reporting*

Adverse events (AE), and serious adverse events (SAE) will be summarised as the number of events and the number of people having events by intervention group and overall.

### *2.6 Scoring of questionnaire outcomes*

See Appendix 1.

### *2.7 Descriptive statistics for outcome measures*

The primary and secondary outcomes will be summarised at baseline and one month and two months post-randomisation by intervention group and overall. Mean and SD or medians and IQR will be used to summarise normally distributed and non-normally distributed variables, respectively.

Where variables have a high proportion of zero values, variables will be summarised (e.g. using medians and IQR) both including and excluding those with a zero value (e.g. app usage data). The percentage of those with non-zero values will also be reported.

## 3. Data analysis plan – Inferential analysis

### *3.1 Main analysis of treatment differences*

All analysis will follow the intention to treat principle as far as possible (Fergusson, Aaron et al. 2002). As all primary and secondary outcomes are continuous, between group difference in mean estimates between Parent Positive and FAU and

associated one-sided 95% confidence intervals (practically a two-sided 90% confidence interval will be computed, with only the limit relating to the one-sided hypothesis reported, e.g. for Parent Positive vs TAU comparisons, the lower limit if a higher score equates to better outcomes, and vice versa) will be reported from the mixed-effects linear analysis of covariance models with repeated measures described below. Two-sided 95% confidence intervals will be presented in the supplementary materials. The significance level will be 5% (one-sided). This significance level will also be used for the secondary outcomes.

### 3.1.1 Analysis of primary outcome

The mean difference in the SDQ conduct scale at one and two-months post-randomisation between Parent Positive and FAU groups will be estimated using a mixed-effects linear analysis of covariance (ANCOVA) model with the repeated one and two month measures as dependent variables and a random intercept at the participant level. We will explore whether adding a random slope over time significantly improves model fit, if so we will include it. The model will include intervention group, time point, intervention group by time point interaction, SDQ conduct problem sub-scale score at baseline, child gender and child age as covariates. If any baseline variables are found to predict missing outcome data as outlined in 3.1.4.5, these will also be included as covariates. The interaction term will be used to extract the intervention effect at one month post-randomisation.

### 3.1.2 Analysis of secondary outcomes

The mean difference between Parent Positive and FAU at two months post-randomisation for SDQ conduct score will be extracted from the model described in Section 3.1.1. The mean differences between Parent Positive and FAU at one month and two months post-randomisation for the rest of the secondary outcomes will be estimated in a similar way to that described in 3.1.1 for the primary outcome, with the same covariates (including any baseline predictors of missing data) except for substituting the relevant baseline measure in each case.

### 3.1.3 Moderation analysis

#### 3.1.3.1 Moderation by baseline variables

Moderation of the Parent Positive versus FAU primary outcome effect by baseline SDQ conduct problems sub-scale score will be explored by exchanging the time by intervention group interaction term in the primary outcome model described in Section 3.1.1 for an intervention group by timepoint by baseline conduct problems score interaction term. We will conclude there is moderation if the overall p-value for this interaction term is $< 0.05$. If so, we will estimate treatment effects for different baseline levels of conduct problems at one month post-randomisation and (the following currently deviates from the protocol, see Section 1.8.5) the conduct outcome at two months post-randomisation.

#### 3.1.3.2 Effect modification by post-randomisation variables

Modelling effects of post-randomisation variables must be conducted carefully and in a principled manner – treatment will likely affect such variables, so they cannot simply be studied as independent variables in models of the treatment effect (Lewis 1999). We will instead examine the effect of app usage on the primary SDQ conduct outcome at one month post-randomisation based on the derived measure of total

time accessing the “Boosters” zone during the one month post-randomisation period (see Sections 1.8.6 and 2.3, app usage variable #6) referred to in the rest of this section as “app usage”). We will use appropriate therapeutic process evaluation measures as outlined in Chapter 3 of the HTA report: *Evaluation and validation of social and psychological markers in randomised trials of complex interventions in mental health: a methodological research programme* by Dunn et al. (Dunn, Emsley et al. 2015). Discrete latent class finite mixture models will be fitted using the structural equation modelling framework, including intervention group, SDQ conduct problem sub-scale score at baseline, child gender and child age as covariates in all models. Due to the complexity of such methods, and having no previous knowledge of the likely extent of app usage/how the app usage variable will be distributed, we will not be able to fully pre-specify this analysis. We will state this transparently in the primary paper. The likely approach will be the following. We will code the total time accessing the “Boosters” zone during the one month post-randomisation period app usage variable in the intervention group into a meaningful binary variable (e.g. 0 = no use, 1 = some use; or 0 = low/no use, 1 = high use), and use principal stratification latent class methods to obtain the mean Parent Positive vs FAU difference in the conduct problem variable for both levels of app usage groups. The binary app usage variable will be coded as necessary in the in the FAU control group to indicate to the software programme that the usage variable is not observed in this group. Within the wider model, the app usage variable will be modelled using logistic regression, with the primary conduct outcome at one month post-randomisation modelled using linear regression. The following currently deviates from the protocol, see Section 1.8.5: we will fit a separate model with the conduct outcome at two month post-randomisation to look at effects on this outcome time point as well. We may explore whether we can fit a principal trajectory model to obtain estimates for both time points from one model. Because the app usage variable is observable in the intervention group, but not in the control group, we need baseline predictors of app usage to get model identification. In addition to the covariates described at the beginning of this section, we will also enter baseline values of the following variables into the model as potential predictors of app usage: DASS score, income, number of adults living in the household, number of children living in the household, full vs. partial or no lockdown, whether parents are working from home or not, and whether children are being homeschooled or not. Note, the Co-SPACE study has been using government dates to designate full/partial/no lockdown rather than the self-report lockdown variables, SPARKLE will do the same. If the models do not run with all predictors, we will need to explore using a subset of these.

If possible, we will apply these methods to any multiply imputed data (see Section 3.1.4.5), however, if this isn’t possible due to the complex nature of these methods, we may extend the models to allow for missing outcomes/latent ignorability as per Dunn et al, 2015. This would entail coding a 0 = complete, 1 = missing variable based on whether the primary conduct outcome is missing or not, and adding this variable to the models described in the previous paragraph. We would then allow this missing data variable to be predicted by all of the above covariates including intervention group. If neither of these approaches is feasible, we may need to restrict these analyses to the complete case population.

### 3.1.4 Statistical considerations

#### 3.1.4.1 Time points

Outcomes are measured at one and two months post-randomisation.

3.1.4.2 Stratification and clustering

Intervention groups will be allocated by simple randomization without blocking or stratification.

The data structure for the outcome is longitudinal with repeated measures over time. The correlation between observations within each participant will be accounted for by using mixed effects linear models with random effects at the participant level as described in Section 3.1.1.

3.1.4.3 Missing items in scales and subscales

The number (%) with complete primary and secondary outcome data at the subscale or scale level will be reported by intervention group and assessment time point and overall.

Please see Appendix 1 for treatment of missing item data.

3.1.4.4 Missing baseline data

The number of participants with missing data at baseline is expected to be minimal based on current Co-SPACE data, and thus, missing baseline data should not represent an issue for the primary analysis. In case of these variables containing missing data, the number with complete data will be summarised and reported. Missing baseline data will be imputed using mean imputation as per White and Thompson's recommendations (White and Thompson 2005).

3.1.4.5 Missing outcome data

One way that missing post-randomisation outcome measures will be dealt with is using maximum likelihood methods to fit the mixed models described in Section 3.1. Such an approach provides valid inferences under the assumption that the data are missing at random. This requires that all variables predicting missing data are included in the models. To assess whether missing outcome data are predicted by baseline variables, we will construct a binary variable coding whether any of the primary or secondary outcomes are missing at either one month or two months post-randomisation. This variable will be the dependent variable in logistic regression models with intervention group, SDQ conduct problem sub-scale score at baseline, child gender and child age as independent variables. Each of the baseline variables that were listed as possible predictors of app usage in Section 3.1.3.2 will be added to this model background in turn as an independent variable and will be considered to predict missing outcomes if there is a significant relationship at a 5% level. Any such variable will be included in the models for analysing the primary and secondary outcomes described in Sections 3.1.1 and 3.1.2.

We will consider multiple imputation (MI) for the primary and secondary outcomes #2-5 in section 1.8.3 only if there are post-randomisation variables that predict missing data in any of these variables and the proportion of observations having any primary or secondary outcome variable missing is greater than or equal to ten per cent (Jakobsen, Gluud et al. 2017, Sullivan, White et al. 2018, Van Buuren 2018). We will assess whether missing outcome data are predicted by post-randomisation variables in a similar way to that described above for assessing baseline variables. The main post-randomisation variable of interest is total time accessing the "Boosters" zone during the one month post-randomisation period app usage variable,

so this is the variable we will assess If this variable predicts missing outcomes, it will be included in the imputation but not analysis models. We will set the random seed number when doing the imputation, so it is reproducible. We will impute separately by intervention group, with an imputation model including all post-randomisation measures of the primary and secondary outcomes #2-5 in section 1.8.3, the total time accessing the Booster zone during the one month post-randomisation period app usage variable, child gender and child age, baseline measures of all outcomes, and all baseline variables found to predict missing outcome data. Including the baseline SDQ conduct measure in the imputation model will cover the need to include any variables that will be explored using an interaction term/as a prespecified moderator (see Section 3.1.3.1), as advised by Sullivan et al and Cro et al (Sullivan, White et al. 2018, Cro, Morris et al. 2020). The number of imputed data sets to be generated will be equal to the proportion of participants with missing data (Harrell Jr 2015). We will endeavour to construct the imputation models as described, but if not viable, we may need to consider a set of simpler imputation models with some variables removed. Where MI is used, we will produce estimates from both the complete case and MI data and include both in the statistical report, with the latter being the main estimates reported in the primary publication. If there are no post-randomisation variables that predict missing outcomes, but there are baseline variables predictive of missing data, we will not do MI and will instead include these baseline variables in the primary and secondary outcome analysis models as described in Section 3.1 and earlier in this section.

#### 3.1.4.6 Method for handling multiple comparisons

No correction will be made for multiple comparisons.

#### 3.1.4.7 Method for handling non-compliance (per protocol analyses)

Since the sample is not restricted to those with a specific clinical need, a per protocol analysis is not planned. However, the effect of app usage in modifying the intervention effect will be examined as described in section 3.1.3.2.

#### 3.1.4.8 Model assumption checks

The linear regression and linear mixed effects models assume normally distributed residuals; this will be checked when describing the data. Residuals will be plotted to check for normality and inspected for outliers. If substantial departures from normality occur, methods allowing for non-normality of residuals will be used (e.g. robust standard errors).

### 3.1.5 Sensitivity analyses

Due to a small proportion of participants having extremely high values for app usage times, sensitivity analysis of the effect modification post-randomisation analysis (described in section 3.1.3.2 above) excluding these outlier values will be performed. We will also summarise app usage data with these outlier values removed. Outliers will be defined as those values more than three IQRs above the upper quartile (generally known as *extreme* outliers).

## *3.2 Planned subgroup analyses*

There are no powered planned subgroup analyses. However, treatment effect moderation by the baseline level of parent reported child conduct problems at one

and two months post randomisation and post-randomisation effect modification by different levels of app usage will be examined as described in section 3.1.3.

### *3.3 Interim analysis*

No interim analysis is planned in this study.

## 4. Software

SPARKLE will be embedded in Co-SPACE, a UK-wide cohort study that utilises the Qualtrics platform. Data will be exported from the platform into Stata file format (.dta). All data processing and statistical analyses for the main trial paper will be performed using Stata versions 15 or higher, with the possible exception of the post-randomisation effect modification analysis described in Section 3.1.3.2, for which we may use structural equation modelling software (i.e., Mplus version 8 or higher).

## SCHEDULE OF ASSESSMENTS AND MEASURES

***Table 1 Data collection points timeline***

| | | | STUDY PERIOD | | | |
|---|---|---|---|---|---|---|
| | | | **Baseline** | **Intervention** | **Follow-up 1** | **Follow-up 2** |
| | | **Timepoint** | **1** | | **2** | **3** |
| | | **Month** | 0 | | 1 | 2 |
| **Sample description** | | | | | | |
| | | Family characteristics and demographics | X | | | |
| **Outcomes** | | | | | | |
| | Child | SDQ conduct problems | X | | X | X |
| | Child | SDQ emotional problems | X | | X | X |
| | Parent | Parental child-related stress and worries | X | | X | X |
| | Parent | Family conflict | X | | X | X |
| | Parent | DASS-21 psychological distress | X | | X | X |
| **Other measures** | | | | | | |
| | | SDQ ADHD symptoms | X | | X | X |
| | | SDQ peer problems | X | | X | X |
| | | SDQ prosocial behaviour | X | | X | X |
| | | CA-SUS service utilisation | | | X | X |
| | | Lockdown circumstances | X | | X | X |
| | | Self-reported adverse events | | | X | X |
| **Interventions randomised 1:1** | | | | | | |
| | | *Parent Positive* | | ← Access to app → | | |
| | | Follow-up As Usual[a] | | ← No access to app → | | |
| **Intervention usage and acceptability** | | | | | | |
| | | Total time spent accessing the *Parenting Boosters* | | | X | X |
| | | Other app usage metrics | | | X | X |
| | | Parent Positive Attitudes | | | X | X |

*Note.* CA-SUS=Child and Adolescent Service Use Schedule; DASS-21=Depression, Anxiety & Stress Scale-21; SDQ=Strengths and Difficulties Questionnaire.

[a]=Parents allocated to Follow-up As Usual will get access to Parent Positive after the Follow-up 2 (T3) assessment.

## REFERENCE LIST


Antony, M. M., P. J. Bieling, B. J. Cox, M. W. Enns and R. P. Swinson (1998). "Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample." Psychological assessment **10**(2): 176.

Cro, S., T. P. Morris, M. G. Kenward and J. R. Carpenter (2020). "Sensitivity analysis for clinical trials with missing continuous outcome data using controlled multiple imputation: a practical guide." Statistics in Medicine **39**(21): 2815-2842.

Dunn, G., R. Emsley, H. Liu, S. Landau, J. Green, I. White and A. Pickles (2015). "Evaluation and validation of social and psychological markers in randomised trials of complex interventions in mental health: a methodological research programme." Health Technology Assessment (Winchester, England) **19**(93): 1-115, v.

Fergusson, D., S. D. Aaron, G. Guyatt and P. Hébert (2002). "Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis." Bmj **325**(7365): 652-654.

Goodman, R. (1997). "The Strengths and Difficulties Questionnaire: a research note." Journal of child psychology and psychiatry **38**(5): 581-586.

Harrell Jr, F. E. (2015). Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis, Springer.

Jakobsen, J. C., C. Gluud, J. Wetterslev and P. Winkel (2017). "When and how should multiple imputation be used for handling missing data in randomised clinical trials–a practical guide with flowcharts." BMC medical research methodology **17**(1): 1-10.

Lewis, J. A. (1999). "Statistical principles for clinical trials (ICH E9): an introductory note on an international guideline." Statistics in medicine **18**(15): 1903-1942.

Machin, D., M. J. Campbell, S. B. Tan and S. H. Tan (2018). Sample sizes for clinical, laboratory and epidemiology studies, John Wiley & Sons.

Moher, D., K. F. Schulz and D. G. Altman (2001). "The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials." Lancet **357**(9263): 1191-1194.

Relton C, Torgerson D, O'Cathain A, Nicholl J. Rethinking pragmatic randomised controlled trials: introducing the "cohort multiple randomised controlled trial" design *BMJ* 2010; 340 :c1066 doi:10.1136/bmj.c1066.

Sullivan, T. R., I. R. White, A. B. Salter, P. Ryan and K. J. Lee (2018). "Should multiple imputation be the method of choice for handling missing data in randomized trials?" Statistical methods in medical research **27**(9): 2610-2626.

Van Buuren, S. (2018). Flexible imputation of missing data, CRC press.

Waite, P. and C. Creswell (2020). The Co-Space Study Report 01: Findings from the first 1500 participants on parent/carer stress and child activity, Oxford, United Kingdom: Univeristy of Oxford.

Waite, P., P. Patalay, B. Moltrecht, E. McElroy and C. Creswell (2020). "Report 02: Covid-19 worries, parent/carer stress and support needs, by child special educational needs and parent/carer work status." Co-SPACE study, UK.

White, I. R. and S. G. Thompson (2005). "Adjusting for partially missing baseline measurements in randomized trials." Stat Med **24**(7): 993-1007.

# APPENDICES

## *Appendix 1 – Scoring of outcome measures*

**SPARKLE Trial Measures and Scoring Document (recruiting from Co-SPACE study)**

This document describes the scoring for the clinical outcomes, household composition and overcrowding, and Parent Positive usefulness score for the SPARKLE trial, which is recruiting from the ongoing Co-SPACE study **http://cospaceoxford.org/.**

**CLINICAL OUTCOMES**

**Strengths and Difficulties Questionnaire (SDQ)**

The Strengths and Difficulties Questionnaire (SDQ) is used to measure the psychological wellbeing of children and young people and is completed by parents/carers and 11–16- year-olds themselves (Goodman, 1997; Goodman, Meltzer, & Bailey, 1998). The SDQ asks about 25 attributes, some positive and others negative, and respondents use a 3-point Likert scale to indicate how much each attribute applies to the child or young person, with 'Somewhat true' always scored as 1 and 'Not True' and 'Certainly True' varying between 0 and 2 depending on the item. See the link below under missing data: the items for this scale have been coded according to this document in the Co-SPACE data, so there is no need for reverse coding after scoring. The 25 items are divided between five scales (each with five items): emotional symptoms, conduct problems, hyperactivity-inattention, peer problems, and pro-social behaviour. Four of the sub-scales (all except pro-social behaviour) are summed to generate a total difficulties score (range: 0-40). There is also an impact supplement that asks whether the respondent thinks that the child or young person has a problem, and if so, inquires further about overall distress, social impairment, burden, and chronicity. This can be used to determine psychiatric 'caseness'. However, the impact supplement is not used to calculate any of the sub-scales. The SDQ has been demonstrated to have high levels of validity and reliability across different countries and settings (Goodman, 2001).

**Scoring:** The 25 items in the SDQ comprise 5 sub-scales of 5 items each. Sub-scales scores are calculated by summing the 5 items, if all of these were completed. The sub-scale scores can range from 0 to 10. Please see Table 1 below for items and variable names.

Only the SDQ conduct and emotional problems subscales are outcomes in SPARKLE, but all items are included in Table 1 below for completeness. The wider SDQ scale will be used as a health economics variable, however, the scoring for this purpose is not covered in this document.

SDQ conduct scale scoring = ptantrum + pobeys* + pfights + plies + psteals
SDQ emotional problems scale scoring = psomatic + pworries + punhappy + pclingy + pafraid
SDQ hyperactivity-inattention scale scoring = prestles + pfidgety + pdistrac + preflect* + pattends*
SDQ peer problems scale scoring = ploner + pfriend* + ppopular* + pbullied + poldbest
SDQ prosocial scale scoring = pconsid + pshares + pcaring + pkind + phelpout
*These items are intrinsically reverse-coded

**Missing data:** The SDQ has directions for scoring where items are missing, further information can be found here:
https://www.ehcap.co.uk/content/sites/ehcap/uploads/NewsDocuments/236/S

DQEnglishUK4-17scoring-1.PDF**.** Each sub-scale score can be scaled up pro-rata if at least 3 items were completed by summing the score across the completed items, multiplying this number by 5, dividing by the number of items completed and rounding this number up to the nearest integer. For example, if 3 items are completed and the sum of the score across these items is 4, the subscale can be scored as (4 x 5)/3 = 6.67, then rounded up to 7.

**Table 1. SDQ items, subscales and Co-SPACE variable names**

| | **SDQ question** | **SDQ subscale** | **Does the question need to be reverse scored? Y/N** | |
|---|---|---|---|---|
| **1** | Considerate of other people's feelings | Prosocial scale | N | pconsid |
| **2** | Restless, overactive, cannot stay still for long | Hyperactivity scale | N | prestles |
| **3** | Often complains of headaches, stomach-aches or sickness | Emotional problems scale | N | psomatic |
| **4** | Shares readily with other children (treats, toys, pencils etc.) | Prosocial scale | N | pshares |
| **5** | Often has temper tantrums or hot tempers | Conduct problems Scale | N | ptantrum |
| **6** | Rather solitary, tends to play alone | Peer problems scale | N | ploner |
| **7** | Generally obedient, usually does what adults request | Conduct problems Scale | Y | pobeys |
| **8** | Many worries, often seems worried | Emotional problems scale | N | pworries |
| **9** | Helpful if someone is hurt, upset or feeling ill | Prosocial scale | N | pcaring |
| **10** | Constantly fidgeting or squirming | Hyperactivity scale | N | pfidgety |
| **11** | Has at least one good friend | Peer problems scale | Y | pfriend |
| **12** | Often fights with other children or bullies them | Conduct problems Scale | N | pfights |
| **13** | Often unhappy, down-hearted or tearful | Emotional problems scale | N | punhappy |

| | | | | |
|---|---|---|---|---|
| **14** | Generally liked by other children | Peer problems scale | Y | ppopular |
| **15** | Easily distracted, concentration wanders | Hyperactivity scale | N | pdistrac |
| **16** | Nervous or clingy in new situations, easily loses confidence | Emotional problems scale | N | pclingy |
| **17** | Kind to younger children | Prosocial scale | N | pkind |
| **18** | Often lies or cheats | Conduct problems Scale | N | plies |
| **19** | Picked on or bullied by other children | Peer problems scale | N | pbullied |
| **20** | Often volunteers to help others (parents, teachers, other children) | Prosocial scale | N | phelpout |
| **21** | Thinks things out before acting | Hyperactivity scale | Y | preflect |
| **22** | Steals from home, school or elsewhere | Conduct problems Scale | N | psteals |
| **23** | Gets on better with adults than with other children | Peer problems scale | N | poldbest |
| **24** | Many fears, easily scared | Emotional problems scale | N | pafraid |
| **25** | Sees tasks through to the end, good attention span | Hyperactivity scale | Y | pattends |

**Depression, Anxiety and Stress Scale (DASS-21)**

The Depression, Anxiety and Stress Scale – 21 items (DASS-21) (Antony, Bieling, Cox, Enns, & Swinson, 1998) is used to measure symptoms of depression, anxiety, stress in parents/carers. The DASS-21 is a short form of Lovibond and Lovibond's (1995) DASS-42. It is a self-report questionnaire consisting of 21 items, with 7 items per sub-scale: depression, anxiety and stress. Items are scored on a scale from 0 ("Never") to 3 ("Almost always"). It has been shown to have good psychometric properties within clinical and non-clinical samples in countries, including the UK (Henry & Crawford, 2005).

**Scoring:** The 21 items in the DASS-21 comprise 3 sub-scales of 7 items each. Scores for each sub-scale are calculated by summing the answers across the 7 items and then multiplying by 2. The latter step is necessary because DASS-21 is the short form of the scale and multiplying by 2 is required to have comparable scores with the full DASS-42. The score for each of the sub-scales can range from 0 to 42 if all items were completed. Further information can be found here: https://www.bristol.ac.uk/media-library/sites/sps/documents/c-change/dass.pdf.

Please see Table 2 below for items and variable names.

The outcome used for SPARKLE will not be the individual subscales, but instead the composite psychological distress scale score from all sub-scales. This is obtained by adding the 3 sub-scales scores together, to provide a scale ranging from 0 - 126.

DASS depression scale scoring = (dass_3 + dass_5 + dass_10 + dass_13 + dass_16 + dass_17 + dass_21) * 2
DASS anxiety scale scoring = (dass_2 + dass_4 + dass_7 + dass_9 + dass_15 + dass_19 + dass_20) * 2
DASS stress scale scoring = (dass_1 + dass_6 + dass_8 + dass_11 + dass_12 + dass_14 + dass_18) * 2
DASS psychological distress composite scale = DASS depression score + DASS anxiety score + DASS stress score

**Missing data:** Qualtrics is designed to prevent missing data for items of DASS-21 and doesn't allow the user to continue until all items are complete. However, each sub-scale score can be scaled up pro-rata if only one item is missing by summing the score across the completed items, multiplying this number by 7, dividing by the number of items completed and rounding this number up to the nearest integer. For example, if 6 items are completed and the sum of the score across these items is 6, the subscale can be scored as (6 x 7)/6 = 7. Further information can be found here: https://www.bristol.ac.uk/media-library/sites/sps/documents/c-change/dass.pdf.

**Table 2. DASS items**

| | DASS question | DASS subscale | |
|---|---|---|---|
| **1** | I found it hard to wind down | Stress | dass_1 |
| **2** | I was aware of dryness of my mouth | Anxiety | dass_2 |
| **3** | I couldn't seem to experience any positive feeling at all | Depression | dass_3 |
| **4** | I experienced breathing difficulty (e.g. excessively rapid breathing, breathlessness in the absence of physical exertion) | Anxiety | dass_4 |
| **5** | I found it difficult to work up the initiative to do things | Depression | dass_5 |
| **6** | I tended to over-react to situations | Stress | dass_6 |
| **7** | I experienced trembling (e.g. in the hands) | Anxiety | dass_7 |
| **8** | I felt that I was using a lot of nervous energy | Stress | dass_8 |

| | | | |
|---|---|---|---|
| **9** | I was worried about situations in which I might panic and make a fool of myself | Anxiety | dass_9 |
| **10** | I felt that I had nothing to look forward to | Depression | dass_10 |
| **11** | I found myself getting agitated | Stress | dass_11 |
| **12** | I found it difficult to relax | Stress | dass_12 |
| **13** | I felt down-hearted and blue | Depression | dass_13 |
| **14** | I was intolerant of anything that kept me from getting on with what I was doing | Stress | dass_14 |
| **15** | I felt I was close to panic | Anxiety | dass_15 |
| **16** | I was unable to become enthusiastic about anything | Depression | dass_16 |
| **17** | I felt I wasn't worth much as a person | Depression | dass_17 |
| **18** | I felt that I was rather touchy | Stress | dass_18 |
| **19** | I was aware of the action of my heart in the absence of physical exertion (e.g. sense of heart rate increase, heart missing a beat) | Anxiety | dass_19 |
| **20** | I felt scared without any good reason | Anxiety | dass_20 |
| **21** | I felt that life was meaningless | Depression | dass_21 |

**Parental child-related stress and worries levels**
This is based on five items routinely measured in Co-SPACE – their children's behaviour (parent_stress_4), wellbeing (parent_stress_5), screen time use (parent_stress_6), education (parent_stress_7) and future (parent_stress_8). "Not at all" is coded as 0, "A little" as 1, "Quite a lot" as 2, "A great deal" as 3. These items form a single scale with an adequate level of internal consistency (alpha > .70) and re-test reliability (r = .71).

**Scoring:** The five questions are summed to generate a total score if all items were completed. The scores range from 0 to 15.

**Missing data:** The scale score can be scaled up pro-rata if at least 3 items were completed by summing the score across the completed items, multiplying this number by 5, dividing by the number of items completed and rounding this number up to the nearest integer. For example, if 3 items are completed and the sum of the score across these items is 4, the subscale can be scored as (4 x 5)/3 = 6.67, then rounded up to 7. The percentage of missing

items allowed to still generate a valid score is greater than the 20% commonly used, however, this is the standard procedure implemented in Co-SPACE.

**Family conflict scale**

This is based on three items relating to arguments between parents, parents and children and siblings routinely measured in Co-SPACE – My child and I argue a lot (num_your_family_3), In my household, there are disagreements between adults about how to parent my child (num_your_family_5), My child and their siblings argue a lot (num_your_family_7). These items form a single outcome measure with acceptable internal consistency (alpha> .54) and test-retest reliability (r = .73). “Not at all" is coded as 0, "A bit" as 1, "A lot" as 2, "Completely" as 3. The score can range from 0 to 9.

**Scoring:** The three responses are summed to generate a total score if all items were completed.

**Missing data:** The scale score can be scaled up pro-rata if at least 2 items were completed by summing the score across the completed items, multiplying this number by 3, dividing by the number of items completed and rounding this number up to the nearest integer. For example, if 2 items are completed and the sum of the score across these items is 5, the subscale can be scored as (5 x 3)/2 = 7.5, then rounded up to 8. The percentage of missing items allowed to still generate a valid score is greater than the 20% commonly used, however, this is the standard procedure implemented in Co-SPACE.

## HOUSEHOLD COMPOSITION AND OVERCROWDING

**Household Composition**

The measures of household composition vary for participants who were recruited to Co-SPACE before and after December 2020. These differences will be outlined below. However, SPARKLE participants should have been asked the post-December 2020 versions of the questions only. Both are provided for completeness.

Reports of household composition are asked such that values of each type of household member (i.e. parent, sibling) currently living within the participant’s home are provided, with summary variables derived as outlined below. It is a self-report questionnaire consisting of between 7 and 15 items (the post- and pre-December 2020 questionnaire versions, see below). Household composition values are recorded at baseline and post mid-March 2021 at follow-up as well. Co-SPACE uses, and SPARKLE will use, baseline household composition variables only. Items are scored on a scale from 0 to 49; the 0 as a possible value for each item (although this can create a problem if the participant does not report on themselves, the procedure has been changed since, see details below in Prior and Post December 2020 sections), and 49 set as an arbitrary maximum in Qualtrics. The total household composition value is coded as ‘total_household’ in the Co-SPACE data and is derived by summing all items across the scale.

**Prior to December 2020.** Participants were asked the following question:

> We'd like to know who lives in your household*. Please tell us how many of each type of person, including yourself. For example, if the child has two brothers, type 2 in the 'Child's brother' box.
>
> *Household = people living in the same house as your child

The 15 items in the prior to Dec 2020 household composition scale are summed and the value of '1' is added to provide a score of household composition. The added value of '1' corresponds to the child being reported about. The total possible score is 735 (i.e. 15 items x max 49 for each item). Impossible scores of 1 or less are replaced with NAs. Given the generous upper threshold for each item, there is no upper threshold which is deemed to be an impossible score, so no other scores are replaced by NAs.
Household composition scoring: total_household = mother + father + stepmother + stepfather + parentpartner + brother + sister + fosterbrother + fostersister + stepsister + stepbrother + grandmother + grandfather + otherrelative + othernonrelative + 1

| | **Pre-December 2020 Household Composition Question** | **Does the question need to be reverse scored? Y/N** | |
|---|---|---|---|
| **1** | Child's mother | N | mother |
| **2** | Child's father | N | father |
| **3** | Child's stepmother | N | stepmother |
| **4** | Child's stepfather | N | stepfather |
| **5** | Parent's partner | N | parentpartner |
| **6** | Child's brother | N | brother |
| **7** | Child's sister | N | sister |
| **8** | Child's foster brother | N | fosterbrother |
| **9** | Child's foster sister | N | fostersister |
| **10** | Child's stepsister | N | stepsister |
| **11** | Child's stepbrother | N | stepbrother |
| **12** | Child's grandmother | N | grandmother |
| **13** | Childs grandfather | N | grandfather |
| **14** | Child's other relative | N | otherrelative |
| **15** | Other non-relative | N | othernonrelative |

**Post December 2020.** Participants were asked the following question:

> We'd like to know who lives in your household*. Please tell us how many of each type of person, EXCLUDING yourself and the child you are answering about. For example, if the child has two brothers, type 2 in the Child's brother box.
>
> *Household = people living in the same house as your child

The 7 items in the household composition scale are summed and the value of '2' is added to provide a score of household composition. The added value of 2 corresponds to the parent taking the study and the child being reported about. The total possible score is 343 (i.e. 7 items x max 49 for each item). Given the changes to the question asking the parent to exclude themselves and the child, the scoring adding 2, and the generous upper threshold for each item, there are no impossible scores that are replaced by NAs.

Household composition scoring: total_household = other_parent + sister_all + brother_all + grandmother + grandfather + otherrelative + othernonrelative + 2

| | **Post-December 2020 Household Composition Question** | **Does the question need to be reverse scored? Y/N** | |
|---|---|---|---|
| **1** | Child's other parent/carer | N | other_parent |
| **2** | Child's sister (including step/foster) | N | sister_all |
| **3** | Child's brother (including step/foster) | N | brother_all |
| **4** | Child's grandmother | N | grandmother |
| **5** | Childs grandfather | N | grandfather |
| **6** | Child's other relative | N | otherrelative |
| **7** | Other non-relative | N | othernonrelative |

**Number of adults and number of children in the household**

Prior to December 2020 participants were asked to list the age of each family member. This is a self-report scale with 15 items. There was no prespecified minimum or maximum range of values.

Allocation into either 'adult' or 'child' group is dictated by condition rather than solely by age (see table below). Then the number of adults and the number of children were calculated separately by summing the number of individuals meeting the two different conditions specified below across the items.

| | **Pre-December 2020 Household Composition Question** | **Does the question need to be reverse scored? Y/N** | **Variable name** | **Condition to be fulfilled to be considered as adult** |
|---|---|---|---|---|
| **1** | Child's mother | N | Age_mother | Include regardless of age |
| **2** | Child's father | N | Age_father | Include regardless of age |
| **3** | Child's stepmother | N | Age_stepmother | Include regardless of age |
| **4** | Child's stepfather | N | Age_stepfather | Include regardless of age |
| **5** | Parent's partner | N | Age_parentpartner | Include regardless of age |
| **6** | Child's brother | N | Age_brother | Exclude regardless of age |
| **7** | Child's sister | N | Age_sister | Exclude regardless of age |
| **8** | Child's foster brother | N | Age_fosterbrother | Exclude regardless of age |

| | | | | |
|---|---|---|---|---|
| **9** | Child's foster sister | N | Age_fostersister | Exclude regardless of age |
| **10** | Child's stepsister | N | Age_stepsister | Exclude regardless of age |
| **11** | Child's stepbrother | N | Age_stepbrother | Exclude regardless of age |
| **12** | Child's grandmother | N | Age_grandmother | Include regardless of age |
| **13** | Childs grandfather | N | Age_grandfather | Include regardless of age |
| **14** | Child's other relative | N | Age_otherrelative | Include only if person is 18 or above |
| **15** | Other non-relative | N | Age_othernonrelative | Include only if person is 18 or above |

**Post December 2020 re adults.** Participants were asked to self-report the number of adults in the household aside from themselves. This is a single item and it is scored on a scale from 1 ("None, I am the only adult in the household") to 7 ("6 or more", coded as 7 because the question is asking about adults aside from themselves). Participants were asked the following question:

| | **Number of Adults** | **Does the question need to be reverse scored? Y/N** | |
|---|---|---|---|
| **1** | Aside from yourself, how many more adults are in your household? | N | adults |

The upper limit for the post December 2020 response is applied to the pre-December 2020 response to make these consistent. There is an additional derived variable called single_adult that identifies households with only one adult.

**Post December 2020 re children.** Participants were asked to self-report the number of children in the household aside from child they are focusing on for the survey. Participants were asked the following question:
Items are scored on a scale from 1 ("None, the child I am answering about is the only child in the household") to 7 ("6 or more", coded as 7 because the question is asking about adults aside from themselves).

| | **Number of Children** | **Does the question need to be reverse scored? Y/N** | |
|---|---|---|---|
| **1** | Aside from the child you are answering about, how many other children are in your household? | N | children_hh |

The upper limit for the post December 2020 response is applied to the pre-December 2020 response to make these consistent.

**Number of Rooms**
The number of rooms index provides details of the number of rooms in the participant's house. It is a self-report questionnaire consisting of 1 item. Items are scored on a scale from 1 ("1") to 15 ("15 or more").

| | **Number of Rooms** | **Does the question need to be reverse scored? Y/N** | |
|---|---|---|---|
| **1** | How many rooms are in your home?<br>• Not including any bathrooms or toilets<br>• If you live in a shared house only count the rooms that are open to you to use<br>If you live in a block of flats, only count rooms in your flat | N | number_of_rooms |

**Overcrowding Index**
The overcrowding index is calculated by dividing the total household composition score (total_household) by the total number of rooms score (number_of_rooms). If the value of the number_of_rooms variable is 15 = 15 or more, the number 15 is used. Values greater than 1 indicate overcrowding.
Overcrowding index scoring: space_avail = total_household / number_of_rooms
This variable is also converted to a categorial overcrowding vs not variable called overcrowded, where overcrowding = overcrowding index >1.

**PARENT POSITIVE EVALUATION QUESTIONS**
This scale will be administered to the parents in the Parent Positive arm only. The questions gathered are included in Appendix A.

The only variables from this scale that will be scored and utilized by the statisticians for the main analysis are called: Booster_Rating, Exchange_Rating, Resources_Rating (see Appendix A). These are each scored from 0 – 6, and will be averaged to derive an overall app usefulness score.

## APPENDIX A. PARENT POSITIVE EVALUATION QUESTIONS

PP_Eval_A

**Your experiences of the Parent Positive App**

We want to know your views of *Parent Positive* app.

Eval_1 Please rate whether you agree with the following:

| | Completely disagree (0) | Somewhat disagree (1) | Neither agree nor disagree (2) | Somewhat agree (3) | Completely agree (4) |
|---|---|---|---|---|---|
| I learnt new information from Parent Positive. (1) | ○ | ○ | ○ | ○ | ○ |
| Using Parent Positive made me reflect on my parenting differently. (4) | ○ | ○ | ○ | ○ | ○ |
| Overall, I found using Parent Positive helpful. (5) | ○ | ○ | ○ | ○ | ○ |
| Parent Positive helped me to deal with a particular parenting challenge I was having. (6) | ○ | ○ | ○ | ○ | ○ |
| I noticed an improvement in my child's behaviour since I started using advice from Parent Positive. (7) | ○ | ○ | ○ | ○ | ○ |
| Parent Positive helped me to feel more confident and happier as a parent. (8) | ○ | ○ | ○ | ○ | ○ |

| | | | | | |
|---|---|---|---|---|---|
| I'd recommend Parent Positive to a friend. (9) | ○ | ○ | ○ | ○ | ○ |
| Parent Positive was easy to use. (10) | ○ | ○ | ○ | ○ | ○ |

PP_Favourite My most favourite thing about *Parent Positive* is:

_____

_____

_____

_____

_____

---

PP_Least_Favourite My least favourite thing about *Parent Positive* is:

_____

_____

_____

_____

_____

---

X→

When_Use_PP When did you use Parent Positive (tick all that apply)?

- ☐ I scheduled regular Parent Positive sessions in my daily routine. (1)
- ☐ I fitted it in whenever I had a few spare minutes. (2)
- ☐ I used it when a particular challenge occurred. (3)
- ☐ When I wanted to connect with other parents and get their advice. (4)
- ☐ If I had a "bad day" with my child. (5)
- ☐ When a new expert Q&A session was posted. (6)

PP_Eval_2 Your views on the Parent Booster zone.

---

X→

Booster_Rating Please provide an overall rating of how useful you found Parenting Skills Booster:

| | Not at all useful (0) | (1) | (2) | Moderately useful (3) | (4) | (5) | Very useful (6) |
|---|---|---|---|---|---|---|---|
| (1) | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

---

PP_Booster_Relevant There are 8 frequent parenting challenges included in Parenting Skills Booster. Please rate each on how relevant its content was to your experience as a parent?

| | Not at all (0) | Slightly (1) | Moderately (2) | Very (3) | Extremely (4) |
|---|---|---|---|---|---|
| Booster 1: Keeping positive and motivated. (1) | ○ | ○ | ○ | ○ | ○ |
| Booster 2: Making sure everyone knows what's expected of them. (2) | ○ | ○ | ○ | ○ | ○ |
| Booster 3: Building your child's self-confidence and trust in you. (3) | ○ | ○ | ○ | ○ | ○ |
| Booster 4: Getting your child to follow instructions. (4) | ○ | ○ | ○ | ○ | ○ |
| Booster 5: Promoting good behaviour. (5) | ○ | ○ | ○ | ○ | ○ |
| Booster 6: How to limit conflict. (6) | ○ | ○ | ○ | ○ | ○ |
| Booster 7: Keeping calm when your kids act up. (7) | ○ | ○ | ○ | ○ | ○ |
| Booster 8: Using sanctions carefully. (8) | ○ | ○ | ○ | ○ | ○ |

PP_Booster_Valuable There are 8 frequent parenting challenges included in Parenting Skills Booster. Please rate each on how valuable its content was to your experience as a parent?

| | Not at all (0) | Slightly (1) | Moderately (2) | Very (3) | Extremely (4) |
|---|---|---|---|---|---|
| Booster 1: Keeping positive and motivated. (1) | ○ | ○ | ○ | ○ | ○ |
| Booster 2: Making sure everyone knows what's expected of them. (2) | ○ | ○ | ○ | ○ | ○ |
| Booster 3: Building your child's self-confidence and trust in you. (3) | ○ | ○ | ○ | ○ | ○ |
| Booster 4: Getting your child to follow instructions. (4) | ○ | ○ | ○ | ○ | ○ |
| Booster 5: Promoting good behaviour. (5) | ○ | ○ | ○ | ○ | ○ |
| Booster 6: How to limit conflict. (6) | ○ | ○ | ○ | ○ | ○ |
| Booster 7: Keeping calm when your kids act up. (7) | ○ | ○ | ○ | ○ | ○ |
| Booster 8: Using sanctions carefully. (8) | ○ | ○ | ○ | ○ | ○ |

PP_Eval_C Your views on the Parenting Exchange zone.

Exchange_Rating Please provide an overall rating of how useful you found the Parenting Exchange zone:

| | Not at all useful (0) | (1) | (2) | Moderately useful (3) | (4) | (5) | Very useful (6) |
|---|---|---|---|---|---|---|---|
| (4) | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

Exchange_Rating_2 Please rate the following statements about Parenting Exchange:

| | Completely disagree (0) | Somewhat disagree (1) | Neither agree nor disagree (2) | Somewhat agree (3) | Completely agree (4) |
|---|---|---|---|---|---|
| I posted on the forum when I needed advice from other parents (1) | ○ | ○ | ○ | ○ | ○ |
| I read through posts to find advice about a parenting issue but never posted myself. (4) | ○ | ○ | ○ | ○ | ○ |
| I used the Exchange to reply to other parents' questions/posts. (5) | ○ | ○ | ○ | ○ | ○ |
| I used the Exchange to connect with other parents socially rather than seek advice. (6) | ○ | ○ | ○ | ○ | ○ |
| It was good to see that I was not the only one finding parenting challenging. (7) | ○ | ○ | ○ | ○ | ○ |
| I only used the Exchange to post questions to experts. (8) | ○ | ○ | ○ | ○ | ○ |

PP_Eval_D Your views on the Parenting Resources.

Resources_Rating Please provide an overall rating of how useful you found Parenting Resources:

| | Not at all useful (0) | (1) | (2) | Moderately useful (3) | (4) | (5) | Very useful (6) |
|---|---|---|---|---|---|---|---|
| (1) | ○ | ○ | ○ | ○ | ○ | ○ | ○ |

Links_Helpful Were there any links that you found especially helpful? Which were they?

_____

_____

_____

_____

_____

End of Block: Parent Positive Evaluation