CEPEO Working Paper No. 25-06

UCL CEPEO and HHCP evaluation of AllChild early intervention programme: 2020-2024

Jake Anders
UCL

Lindsey Macmillan
UCL

Gill Wyness
UCL

Yana Manyukhina
UCL

Dominic Wyse
UCL

Over the past four and a half years, UCL Centre for Education Policy & Equalising Opportunities (CEPEO) and the Helen Hamlyn Centre for Pedagogy (0-11 Years) (HHCP) have been working closely with AllChild to design and carry out impact and process evaluations of AllChild's impact programme. Our primary impact evaluation design is statistical matching design based on estimating young people's propensity of being selected for the AllChild programme. In addition, we seek to come closer to estimating the causal impact of AllChild participation using a discontinuity-based design that also exploits a feature of these selection processes. The implementation and process analysis conducted observations, interviews and informal conversations with a range of informants involved in delivery or receipt of the AllChild programme, including case study participants, followed by thematic analysis of the qualitative data collected. This report documents the analysis and findings of the work during the project.

VERSION: July 2025

Suggested citation: Anders, J., Macmillan, L., Wyness, G., Manyukhina, Y., & Wyse, D. (2025). *UCL CEPEO and HHCP evaluation of AllChild early intervention programme: 2020-2024* (CEPEO Working Paper No. 25-06). UCL Centre for Education Policy and Equalising Opportunities. https://EconPapers.repec.org/RePEc:ucl:cepeow:25-06.

Disclaimer

Any opinions expressed here are those of the author(s) and not those of UCL. Research published in this series may include views on policy, but the university itself takes no institutional policy positions.

CEPEO Working Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

Highlights

- We were unable to identify consistent evidence of impact associated with participating in the AllChild programme compared to the estimated counterfactual on the available quantitative measures. However, we should temper these findings in a couple of ways.
- First, there were severe challenges in achieving a sufficiently similar comparison group to AllChild participants. The lack of positive estimates may be caused by this issue, especially as the estimates from our discontinuity design — which we would anticipate being closer to causal estimates — tell a more positive picture. However, the estimates from the discontinuity design, while quite substantial, are imprecise: even with larger estimates they are not statistically distinguishable from no impact, and are limited to primary schools.
- Second, there are positive spots evident. These include encouraging findings on domains of the SDQ and findings that suggest there may have been more positive impacts where the programme is delivered in full, for example, more positive impacts on attendance evident in both primary and secondary contexts when focussing only on participants who received AllChild's core commitment.
- Such findings were consistent with qualitative implementation and process evaluation work, which found that the programme demonstrates a strong emphasis on socio-emotional support, aligning well with schools' primary motivation for enrolment, and success in improving recipients' socio-emotional well-being. As such, we caution that this report should not be interpreted as strong evidence of no impact, but rather lack of strong evidence in favour of impact on these particular measures.

Why does this matter?

It is important, but not always easy, to provide rigorous evidence of the impact of targeted intervention programmes. We would encourage prioritising an evaluation design that delivers a well-balanced comparison group even if this means less flexibility in other aspects of the design.

UCL CEPEO and HHCP evaluation of AllChild early intervention programme: 2020-2024

Jake Anders
Lindsey Macmillan
Gill Wyness
Yana Manyukhina
Dominic Wyse

Table of contents

Executive Summary

Over the past four and a half years, UCL Centre for Education Policy & Equalising Opportunities (CEPEO) and the Helen Hamlyn Centre for Pedagogy (0-11 Years) (HHCP) have been working closely with AllChild to design and carry out impact and process evaluations of AllChild's impact programme.

Our primary impact evaluation design is statistical matching design based on estimating young people's propensity of being selected for the AllChild programme, with the design developed explicitly to approximate AllChild's selection meetings by using the quantitative variables available as part of those meetings. In addition, we seek to come closer to estimating the causal impact of AllChild participation using a discontinuity-based design that also exploits a feature of these selection processes. The implementation and process analysis conducted observations, interviews and informal conversations with a range of informants involved in delivery or receipt of the AllChild programme, including case study participants, followed by thematic analysis of the qualitative data collected.

While we have been able to identify comparison groups that are a substantial improvement on attempting to evaluate the impact of AllChild simply by comparing participants with non-participants, we caution that observable (and, likely unobservable) differences do remain between the treatment and comparison groups. This is the case even where we have taken quite an aggressive approach to identifying well-matched comparison groups at the expense of sample size – ultimately young people selected into AllChild are extremely different to their peer group and, as a result, it is extremely difficult to identify a suitable number of truly comparable individuals as a comparison group. Consequently, we are unlikely to be able to recover impact estimates that impartial observers would consider to be truly causal. This is particularly the case because the comparison groups assembled generally have average characteristics that are correlated with stronger outcomes (educational and wider) than the AllChild participants to whom they are compared.

We were unable to identify consistent evidence of impact associated with participating in the AllChild programme compared to the estimated counterfactual on the available quantitative measures. However, we should temper all of our quantitative findings in a couple of ways:

- First, the challenges noted in achieving a sufficiently similar comparison group to AllChild participants. The lack of positive estimates may be caused by this issue, especially as the estimates from our discontinuity design — which we would anticipate being closer to causal estimates — tell a more positive picture. However, the estimates from the discontinuity design, while quite substantial, are imprecise: even with larger estimates they are not statistically distinguishable from no impact, and are limited to primary schools.

- Second, there are positive spots evident. These include encouraging findings on domains of the SDQ and findings that suggest there may have been more positive impacts where the programme is delivered in full, for example, more positive impacts on attendance evident in both primary and secondary contexts when focussing only on participants who received AllChild's core commitment. Such findings were consistent with qualitative implementation and process evaluation work, which found that the programme demonstrates a strong emphasis on socio-emotional support, aligning well with schools' primary motivation for enrolment, and success in improving recipients' socio-emotional well-being.

As such, we caution that this report should not be interpreted as strong evidence of no impact, but rather lack of strong evidence in favour of impact on these particular measures. Challenges inherent in the evaluation design that was jointly deemed feasible (chosen to include a number of features valuable to addressing a number of additional research questions but, ultimately, negatively affecting the overall design) have contributed to this outcome, along with attrition rates (i.e., missing outcome measures, especially for endpoint outcomes) to a much greater extent than is typically needed for robust evaluation evidence. We would encourage AllChild and other organisations seeking to continue their evaluation journey to plan to ensure outcomes data can be collected from as many participants and members of the comparison group as possible and prioritise an evaluation design that delivers a well-balanced comparison group even if this means less flexibility in other aspects of the design.

1 Introduction

Over the past four and a half years, UCL Centre for Education Policy & Equalising Opportunities (CEPEO) and the Helen Hamlyn Centre for Pedagogy (0-11 Years) (HHCP) have been working closely with AllChild to design and carry out impact and process evaluations of AllChild's impact programme.

UCL was appointed by AllChild (known at the time as West London Zone; WLZ) in 2020 to act as a partner on its evaluation journey. As such, our work was not simply to collect data and analyse data for the researcher purposes as we saw them, but also to work with AllChild to support the monitoring and evaluation decisions they were taking locally and respond to their organisational evaluation needs through the course of the partnership. We have sought to balance this with conducting as rigorous and independent an evaluation as possible, agreeing at the outset strong principles in support of this aim, such as agreeing and sticking to impact evaluation plans and the importance of publishing our findings.

UCL worked with AllChild across four evaluation cohorts that began in consecutive academic years — 2020/21, 2021/22, 2022/23 and 2023/24 — with baseline data collected in the autumn term of each of these years, midpoint data collected in the summer term of each of these years, and endpoint data collected in the summer term of the subsequent academic year (with the exception of the final evaluation cohort).

In the course of the project, UCL has carried out impact and process evaluation work on each of these individual cohorts, which have been provided to AllChild, while this report describes analyses and findings based on data pooled together from all four cohorts.

1.1 AllChild Impact Programme

AllChild's Impact Programme is a targeted early intervention programme that offers intensive and comprehensive support to the identified participants, tailored to their individual needs. AllChild works with both primary and secondary schools to identify (using a highly data-informed approach) children and young people aged from 3 to 16 who would benefit from its programme, subsequently bringing together families, schools, charities and local government in order to delivery this package of support over a period of two years.

AllChild designed the programme to focus on early intervention and, as such, do not target children and young people who require more intensive support than the programme can provide, for example because they have very low levels of school attendance (especially when combined with a need for academic support), or those with an Education Health and Care Plan (EHCP).

The programme is especially coordinated through 'Link Workers' who are employed by AllChild but based in partner schools. The Link Worker works with participating children and young people as a 'trusted adult' who guides them through the two years of the programme, coordinating the participant's individual support plan (which is drawn up in

consultation with the child, their parents and teachers) with much of the specialist support sessions then commissioned from other charities. AllChild see the Link Workers as having a central role in ensuring the right support at the right time for each individual.

As well as Link Workers, all schools participating in the AllChild programme identify an In-school Lead, who plays a vital strategic role in the programme's implementation within schools. This position is typically held by a senior staff member, usually the head teacher or safeguarding lead. The In-school Lead's primary responsibilities encompass strategic oversight and communication, including:

- Initial identification of children who could benefit from the programme
- Active participation in the final selection of programme participants
- Regular liaison with Link Workers regarding participant progress
- Receipt and review of progress reports about participating children
- Strategic oversight of safeguarding and child protection matters

Drawing on their comprehensive knowledge of the school population, In-school Leads compile initial lists of children who might benefit from the programme. They then work collaboratively with AllChild to refine these selections, ensuring the programme reaches those who would benefit most from its support. In-school Leads maintain regular communication with Link Workers through scheduled update meetings, review of progress reports, ad-hoc consultations about emerging issues or concerns. A key benefit consistently highlighted by In-school Leads is their intentional separation from operational and logistical responsibilities. Unlike many other intervention programmes, AllChild's model explicitly assigns all organisational duties to Link Workers, allowing In-school Leads to focus on strategic oversight without the burden of day-to-day programme management.

The programme has strong roots in West London (although it is now seeking to build such roots and adapt its programme appropriately for other areas of need) and is, as such, a place-based approach designed with and by the local community. Particularly in this sense, the programme has parallels with area-based programmes evaluated in the US, such as the Quantum Opportunity Program (QOP) and the Harlem Children's Zone (HCZ), which share similar approaches and goals.

QOP is a mentoring programme that provides after-school supplemental academic education, developmental activities, and community service to low performing high-school students throughout their four years of high school. Evidence from a randomised trial shows that participants obtained their high school diplomas earlier and were more likely to attend postsecondary schools than non-participants. This programme was particularly successful among youths with ex-ante high-predicted risk of drug-use.

HCZ provides children from a particularly deprived area in New York City with holistic support, combining strong academic support with community services designed to ensure the environment outside of school is also positive for children. Assignment to the programme is randomised when schools participating in the programme are

oversubscribed, with participants selected by a lottery in such circumstance. Researchers exploit this feature in order to identify the causal impact of the programme, finding that it is very effective at increasing the achievement of disadvantaged children: this programme alone is enough to close the black-white achievement gap in mathematics by the end of middle school.

As is evident from the examples above, often such programmes have been evaluated using randomised controlled trials as a way of assembling an appropriate comparison group. Such an approach was ruled out by AllChild as not compatible with their organisational aims. As such, our aim has been to design and conduct a quasi-experimental impact evaluation, identifying appropriate comparison groups from the rest of the cohort from which AllChild's participants are drawn, while still allowing for appropriate causal inference insofar as this is possible within these constraints.

1.2 Aims of evaluation

As noted above, UCL were appointed to this programme as partners on AllChild's evaluation journey. As such, the aims of the project have evolved over the course of the four-and-a-half-year project. The aims agreed at the outset were encapsulated in the following five research questions:

1. What is the impact of the programme on short term (end of programme) outcomes for children and young people over the 2 years of the programme?
2. What is the impact of the programme on long term (end of setting) outcomes for children and young people after they have completed the 2 year programme?
3. What is the impact of the programme on the wider community - schools, families, community?
4. Was the programme implemented as designed in the Theory of Change? How did this influence the impact on short term and long term outcomes for children and young people?
5. What is the cost-benefit of the programme? Is the programme value for money?

The core of these aims, seeking to identify impacts using both quantitative and qualitative analyses of data either provided by AllChild or collected by UCL in the course of the project, has not changed. However, there has been variation in the extent of our ability to deliver aspects depending upon the availability of data to do so. Furthermore, based on our findings, we have not proceeded to cost-benefit analysis of the findings.

When approaching each aim, we have not sought simply to identify each as being approached either by quantitative impact evaluation approaches or qualitative process evaluation approaches. Rather, each aim has been considered as part of work from both of these perspectives, which we then seek to bring together as appropriate in this report. Nevertheless, it is of course the case that some aims lend themselves to one of these approaches more than the other, due to limitations of data availability or the nature of the question itself, leading to differences of emphases where appropriate.

While we have successfully been able to identify comparison groups that are a substantial improvement on attempting to evaluate the impact of AllChild simply by comparing participants with non-participants, we caution that observable (and, likely unobservable) differences do remain between the treatment and comparison groups. This is the case even where we have taken quite an aggressive approach to identifying well-matched comparison groups at the expense of sample size – ultimately young people selected into AllChild are extremely different to their peer group and, as a result, it is extremely difficult to identify a suitable number of truly comparable individuals as a comparison group.

Consequently, we are unlikely to be able to recover impact estimates that impartial observers would consider to be truly causal, but still provide evidence that we judge likely to be persuasive. This is particularly the case because the comparison groups assembled generally have average characteristics that are correlated with stronger outcomes (educational and wider) than the AllChild participants to whom they will be compared. Hence, the directional bias we might expect on our estimates of treatment would be for an underestimate, not an overestimate, of treatment. Any negative estimates would need to be interpreted with this in mind, while we judge that any positive estimates would be unlikely to be undermined by bias in the treatment estimates.

1.3 Outline of report

The report now proceeds as follows. We begin by describing the methods used in the course of this project in Section 2, before setting the data collected and used and, hence, describing the cohort who are selected for participation in the AllChild early intervention in Section 3. Building on these, we report our main findings in Section 4 and unpack these to attempt to understand potential reasons for these findings in Section 5. Finally, we conclude in Section 6.

2 Evaluation methods

In this section, we set out the methods used to evaluate the AllChild early intervention programme. We use multiple approaches to addressing the project's aims, spanning both quantitative impact evaluation methods and primarily qualitative process evaluation methods.

Our primary impact evaluation design is statistical matching design based on estimating young people's propensity of being selected for the AllChild programme, with the design developed explicitly to approximate AllChild's selection meetings by using the quantitative variables available as part of those meetings. In addition, we seek to come closer to estimating the causal impact of AllChild participation using a discontinuity-based design that also exploits a feature of these selection processes. The implementation and process analysis conducted observations, interviews and informal conversations with a range of informants involved in delivery or receipt of the AllChild programme, including case study participants, followed by thematic analysis of the qualitative data collected. We now describe each of these designs in further detail.

2.1 Impact evaluation: matching design

Our primary impact evaluation method was developed in discussion with AllChild to select an approach that could meet multiple aims that are important to their organisation's process of evaluation, including the need to provide results each year during the project (which also increased challenges around sample size for these year-by-year analyses), differences in data collection processes between participating schools, and a desire to explore differences in estimates associated with variation in programme receipt.

To this end, we selected a matching design that is explicitly built upon attempting to approximate the selection process used by AllChild to identify programme participants using the administrative data available to us. More specifically, we attempt to approximate the selection process using a binary choice regression model of the following form:

$$Pr\ (Select) = \alpha + \beta' X_{ijt} + \varepsilon_{ijt}$$

where X is a set of pupil characteristics including year group, baseline attendance and baseline attainment (discussed further below) for pupil i in school j in year t.

We estimated a separate selection model in each school in the cohort for three key reasons:

1. This tracks our approach of attempting to mirror, as far as feasible, the actual approach taken by AllChild and schools, i.e., a separate selection exercise in each school;
2. Relatedly, it's possible that the relationship between pupil characteristics and the probability of selection varies from school to school, which a combined model would ignore.

3. There are some differences between schools in the pupil characteristics available at baseline (e.g., different measures of baseline attainment);

We note that there are potential drawbacks of this approach, however, in that the selection models will each be estimated on smaller samples leading to less explained variance, attenuation bias in the estimates and, hence, noisier estimates of the probability of selection obtained from these models. However, we judge these to be minor compared to the advantages that lead us to use separate models.

Following naturally from the use of school-specific selection models, we then carried out school-specific matching exercises, selecting matches for each child selected onto the programme based on an optimal "nearest neighbour" matching algorithm [@Gu1993] in terms of their estimated selection probability, while enforcing exact matching on year group. Using within-school and within-year group matching ensured that participants are compared with non-participants with important shared environment characteristics, while use of optimal matching would be expected to improve balance when there is a relatively restricted pool of potential comparators.

Using the approach outlined above, we carried out two matching exercises for each school in each year. One to identify a 'wide' matched sample (choosing two "nearest neighbours" for each participant based on the estimated selection probabilities estimated) and one to identify a 'narrow' matched sample (choosing only one "nearest neighbour" for each participant and excluding matched — and the relevant participant — if these did not meet a minimum match quality threshold). For avoidance of doubt, "nearest neighbours" refer to statistical similarity on estimated selection probability, rather than anything to do with geographical proximity.

The 'wide' matched samples have more remaining imbalance, which does raise more questions about bias in the treatment estimates, especially as the magnitudes of that imbalance are often large to the point that we doubt the ability to deal with them as part of our analysis method outlined below. The 'narrow' samples have relatively modest imbalance, but are on a much smaller sample size which a) will have reduced statistical power as a result and b) likely reflects a non-random sub-set of AllChild participants making the impact estimate harder to generalise. We recognise that having both reported below does have disadvantages in terms of not have a single 'right' answer – but that would not really be the case if we just picked one anyway. We think this is more than outweighed by the advantages of carrying out impact estimation with both samples and seeking to understand any differences in results.

We then identified these comparison pupils (along with others identified for the discontinuity design, discussed below) to AllChild so that they could collect outcomes data for this comparison group as well as programme participants. This was done at the end of the academic year in which selection was carried out (as 'midpoint' outcomes) and at the end of the subsequent academic year (as 'endpoint' analyses, reflecting the designed duration of an AllChild programme).

After collection of the outcomes (either midpoint or endpoint), main impact analyses were then carried out on an 'intention to treat' basis, which means that we aimed to include all members of the identified participant and comparison group in our analyses regardless of the extent to which they have received delivery of the AllChild programme over the course of the academic year. This is consistent with evaluation best practice and supports interpretation of the impact of a participant being selected onto the AllChild programme (as distinct from receiving the full programme), which is typically held to be a more realistic impact estimation given that, in most real life situations, a proportion of individuals who were selected onto any programme do not go on to receive it (in full). However, this aim must be tempered by the reality that where AllChild participants stopped participating in the programme, it was also quite likely that it would not be possible to collect data about their outcomes.

Note that in all cases, where a participant's outcomes data are missing the relevant matched comparison pupils' data have also be removed from the analysis sample. Where all matched comparison pupils' data for a given participant are missing, the relevant participant's data has also been removed from the analysis sample. No substitutions of alternative comparison pupils occur in such circumstances. Sample attrition (which is exclusively because of these reasons) is reported in order to provide appropriate context, since attrition may be expected to bias impact estimates and so high levels of attrition may be considered to reduce the credibility of impact estimates. For avoidance of doubt, we highlight that this is 'attrition' in the sense of having lost a participant from the evaluation cohort – not necessarily that they have dropped out of the programme. This does not in itself affect whether we include someone in the analysis sample, which we will do as long as their data are available, although as noted above AllChild indicate that it is rare that they are able to collect data from those who stop participating.

We note that imbalance in observable characteristics remains in our matched samples – particularly the 'wide' matched sample, but also to a more moderate degree in the 'narrow' matched sample. Given this, we include baseline characteristics in our analysis models to adjust for these in minimising bias in our treatment estimates. This is in line with established literature on adjustment within matched samples [@Caliendo2008].

Consequently, we estimate the impact of participating in the AllChild early intervention programme using a regression model of the following form estimated on the matched samples as constructed above:

$$Y_{ijt} = \alpha + \gamma AllChild_{ijt} + X_{ijt} + \zeta + \varepsilon_{ijt}$$

where Y is our outcome of interest for pupil i in school j in year t, AllChild is an indicator of whether the pupil is part of the AllChild early intervention programme cohort, X is a set of covariates (discussed further below), ζ is a set of covariates to adjust for structural features of the evaluation design (including school and evaluation year), and γ recovers our estimate of the impact of interest. ε is the idiosyncratic error term, with standard errors also calculated taking into account the structural features of the matching process (i.e.,

clustering between participants and the specific non-participants with whom they have been matched).

For simplicity and transparency, we keep the set of covariates included as consistent as possible, and these are informed by imbalance in our matched samples at baseline: pupil premium eligibility, whether the individual is male, the pupils' wellbeing rank, attendance measure, SDQ total score, attainment measure, whether the individual has any special educational needs, and their confidence score. Many of these were used as part of the matching process and, so, there is no missing data. Where these is some missing data (almost exclusively for SDQ total score, as this was collected after matching was carried out) we use a simple mean imputation and missing indicator strategy to avoid having to exclude an individual's data for this reason.

The form of the outcome measure of interest in a given analysis will inform the interpretation and any appropriate adjustment to the interpretation of γ. If we have a continuous outcome, we convert this into a Cohen's d effect size for interpretability and comparability with other impact evaluations (although we caution against comparing with effect sizes from RCTs, for example, as these are typically lower than for quasi-experimental evaluations). Where we are looking at binary variables (proportions of pupils who are considered at risk based on AllChild's risk thresholds) we also report a percentage change for ease of interpretation.

We also look to estimate impacts of AllChild among sub-groups of the sample. In such cases, the impact analysis is repeated on the relevant sub-group of participants and the specific group of individuals with whom they have been matched (i.e., not among the wider comparison group). However, we highlight at this point limitations on our ability to do this across all potential groups of interest. Given sub-group analysis inevitably reduces sample sizes of our analysis, we only carry out such analysis in our pooled sample of primary schools.

Specifically, we estimate impacts among sub-groups defined by splitting the sample of programme participants into three equally sized groups depending upon the total number of programme hours they received and comparing each of these groups of participants with their relevant matched comparators (note, however, that the sample size of these analyses are not equal because there are more matched non-participants among the comparison sample).

2.2 Impact evaluation: discontinuity design

At the outset of the project, it was agreed that, in addition to the core matching evaluation approach as described above, data would also be collected to allow estimation of impact using a discontinuity design. By discontinuity design, we mean exploring the impact of the AllChild programme by comparing individuals who sit either side of cut-offs in continuous criteria used to inform the cohort selection processes which, hence, appear to substantially change young people's probability of selection onto the programme despite very similar levels of the continuous criterion itself.

This description is rather abstract in isolation. What does it mean in practice? Young people's attendance rate at baseline is used to inform cohort selection decisions. In particular, whether or not young people are above or below particular cut-offs appeared (based on analysis of existing data at time of design) to shift young people's probability of selection by a substantial amount. As such, we planned to compare young people on either side of (but close to) these cut-offs.

Our first step was to check a) how many young people fall close to the discontinuity threshold and b) the extent to which young people close to the threshold on either side are, indeed, similar to one another, except in respect of their attendance. Both of these criteria require a definition of 'close to the threshold'.

But how close to the threshold we are affects these two factors in opposite directions: the closer to the threshold we restrict our sample the fewer young people there will be on either side, but the more likely it is that they will be similar to one another. Looked at from the other direction, as we widen the window (or 'bandwidth') around the discontinuity, we will get larger sample sizes but the characteristics of those on either side of the threshold are likely to start differing given the correlation between attendance rates and young people's other characteristics. As such, we must find some compromise definition of 'close to the threshold' and this is a central challenge of this kind of design.

So, using the same methods as those used to report balance in the matching design sample (i.e., means and standardised differences), we explored balance between those on either side of the attendance threshold as we change the size of the window around this. We were also mindful of how the sample size changed as we did this. Based on these explorations, we report below on a sample in which we have obtained good balance across the baseline characteristics available while maintaining as large a sample size as possible.

If our research question of interest was how much being just above the attendance threshold increased young people's probability of selection into the AllChild cohort then this would be all we needed to do at this point. But, for this work, this is only an interim question that serves our attempt to answer the question of whether selection into the AllChild cohort affects young people's academic and wider outcomes. Alternatively, if young people's probability of selection into the AllChild cohort on either side of this threshold then, similarly, this would be all that we needed to know. However, in reality, young people's position relative to attendance rate cut-offs is, obviously, not the only factor that affects their probability of cohort selection.

As a result, we are using what is known as a 'fuzzy' discontinuity design, in which we need to take an additional step to extrapolate from how much young people's probability of selection changes at the threshold to our actual research question about the impact of participating in the programme. This means we need to know how much young people's probability of participation does, indeed, change at the threshold. As a result, we will report on this difference among our proposed discontinuity sample.

This is done both with the simple difference in cohort selection probability between the two groups, but also using a statistical test known as an F test, which answers this question in a slightly different way in terms of strength of being above the threshold as a predictor of cohort selection. The statistic obtained from an F test is useful for various diagnostic purposes, including identifying the risk of the difference in selection probability around the discontinuity being a 'weak' instrument that could attenuate our impact estimates.

With those preliminaries completed, estimation of the impact using this discontinuity design is carried out using two stage least squares regression models on the relevant samples identified above. The first stage of these models participation in AllChild as a function of attendance relative to discontinuity:

$$AllChild_{ijt} = \alpha + \gamma AttendDiscont_{ijt} + \mu_{ijt}$$

where AllChild is an indicator of whether the pupil is part of the AllChild early intervention programme cohort for pupil i in school j, AttendDiscont is a binary indicator of whether or not an individual is above or below the relevant attendance discontinuity, and μ is the idiosyncratic error term, with standard errors also calculated taking into account the structural features of the matching process (i.e., clustering between participants and the specific non-participants with whom they have been matched). The predicted participation status from this model $\widehat{AllChild}$ is then used in the second stage impact estimation model (which is otherwise similar to the models used for the matching design):

$$Y_{ijt} = \alpha + \gamma \widehat{AllChild}_{ijt} + \zeta + \varepsilon_{ijt}$$

where Y is our outcome of interest for pupil i in school j, γ recovers our estimate of the impact of interest and ε is the idiosyncratic error term, with standard errors also calculated taking into account the structural features of the design.

2.3 Process evaluation: qualitative design

The implementation and process evaluation (IPE) aimed to enhance understanding of the short-term and long-term outcomes of the AllChild programme on children and young people (CYP) and the wider community through in-depth qualitative work. A focus on CYP's agency was a significant aspect of the IPE, with particular attention given to how agency shaped CYP's experiences and feelings about the programme. For the purposes of the evaluation, we defined agency through the concept of choice.

The IPE engaged directly with:

- CYP receiving the programme to gather pupils' perspectives on their experience with the programme.
- Individuals working with the CYP in other capacities, such as head teachers, teachers, and parents, to assess the impact on the wider community.

- Those delivering the programme, including Link Workers and professionals involved in partner activities, to evaluate how the programme was implemented.

The research methods for the IPE included, in each school:

- Termly observations of partner activities and other relevant interactions involving case study pupils (CSPs) – one per term.
- Termly individual interviews with CSPs – one per term.
- Individual interviews with Link Workers and In-school Leads –two per year, at the start and at the end of the academic year.
- Observations three cohort selection meetings to understand the process of selection of programme participants and factors influencing this process.
- An observation of a partner allocation meeting to understand the process of allocating partner activities to CYP and factors influencing this process.
- Informal conversations with CSPs and adults, as appropriate.
- Post-programme interviews with CSPs from Cohort 1 to gain insights into their learning experiences and lives after the programme had been completed. The original plan included five visits (one for each CSP); however, we were only able to interview two former CSPs - one in primary school and one in secondary school - due to difficulties in securing access to the CSPs.

The analysis followed a thematic approach involving the following steps:

1. Full transcription of the interview recordings.
2. Reading and re-reading the transcripts.
3. Making analytic notes on transcripts to establish analysis codes.
4. Initial coding: grouping data into themes via comments in Microsoft Word documents.
5. Further coding and recoding, refining themes.
6. Final re-reading and checking of codes and themes (see Appendix).

3 Data and description of AllChild cohort

In this section, we discuss the data collected directly by UCL researchers in the course of the process evaluation, as well as the administrative data made available to us by AllChild for the purposes of this evaluation.

3.1 Ethics and data protection

This research project was approved through the approval processes of UCL Institute of Education's research ethics committee and data processing for the project was approved under UCL's data protection procedures. Data processing is carried out by UCL for research purposes under UCL's statement of tasks in the public interest as set out by its statute and charters. Information on the research and an opportunity to object to UCL's processing of data was provided to children and parents in the schools where AllChild works and no data was shared with UCL where objections were received; further opt-in consent was sought among participants in any process evaluation activities.

A data sharing agreement was put in place between AllChild and UCL to cover the steps taken by both parties to ensure the security of transfer and processing. Furthermore, data were shared without any direct identifiers that would allow UCL to identify an individual. An AllChild-provided pseudonymous identifier was included for purposes such as to allow linking individuals' data over time but this was meaningless from UCL's perspective under the principle of data minimisation and privacy by design.

3.2 Qualitative data collected as part of implementation and process evaluation

Over the four years of the IPE research, 10 pupils were selected as the focus of in-depth case studies aimed at providing direct insight into the experiences and views of the CYP receiving the AllChild intervention. The case studies were distributed across the evaluation as follows:

- IPE Cohort 1: 5 case study pupils in years 1-2 of the evaluation (2020-2022)
- IPE Cohort 2: 5 case study pupils in years 3-4 of the evaluation (2022-2024)

All case study pupils were selected in consultation with AllChild. The key selection criteria included the pupils' school year, gender, SEN status, Pupil Premium receipt, identification by AllChild as part of the target population, and previous participation in the AllChild programme. These criteria were used to ensure appropriate breadth in the sample. Table 1 details the sample of case study pupils (CSPs).

Table 1: Process Evaluation Case Study Pupils (CSPs)

CSP and Cohort years	Year	Gender	Pupil Premium	SEN	AllChild target	Prior AllChild experience
CSP1 2021	7	Male	Yes	No	Yes	No

CSP2 2021-2022	8	Male	No	No	Yes	No
CSP3 2021-2022	5	Male	No	No	Yes	No
CSP4 2021-2022	3	Female	Yes	Yes	Yes	No
CSP5 2021-2022	4	Female	Yes	Yes	Yes	No
CSP6 2023-2024	5	Male	No	No	Yes	No
CSP7 2023-2024	4	Male	Yes	Yes	Yes	No
CSP8 2024	3	Female	Yes	Yes	No	No
CSP9 2023-2024	2	Female	Yes	Yes	Yes	No
CSP10 2023	8	Female	No	No	No	No

Over the four years of the IPE, some of the initially chosen CSPs dropped out due to changes in personal circumstances. CSP1 left the school halfway through the programme. In consultation with AllChild, it was decided not to recruit a new CSP and instead focus more intensively on the remaining four CSPs, extending each case study to include Interviews after the end of the programme. CSP3 was selected as a case study pupil in May 2021 after it became clear that the initially chosen pupil would be leaving the school in the summer. Similarly, CSP7 and CSP8 were selected at later stages to replace other CSPs who had dropped out during the evaluation. CSP10 did not continue as a CSP in the final year because the school decided to discontinue the AllChild programme.

All new CSPs were identified in consultation with the respective Link Workers, ensuring a demographic profile similar to the originally selected pupils. This approach helped to maintain consistency against the selection criteria in the evaluation, despite the changes in participant composition.

The following data sets were generated:

1. Audio and video recordings of interviews with adults and CSPs (video applied where interviews were conducted online).
2. Field notes of observations of partner activities and interactions between the Link Workers and the CSPs.
3. Other field notes.

These data provided the basis for the analysis that informs our findings for the implementation and process evaluation.

3.3 Administrative data provided by AllChild for quantitative analysis

The impact evaluation uses data collected by AllChild in the course of their programme administration, shared for the purpose of this research as described above. AllChild shared data across fourth evaluation cohorts, those of academic years 2020-21 (cohort 1), 2021-22 (cohort 2), 2022-23 (cohort 3) and 2023-24 (cohort 4). For the purposes of this report, we are typically reporting analyses based on pooling the data from across all available cohorts for the analyses that we are doing. For baseline and mid-point outcome analyses this means all four cohorts; for end-point outcome analyses this means the first

three of these, since end-point outcome data is collected at the end of the subsequent academic year, so is not yet available for cohort 4 at time of writing.

Data for each full cohort of young people considered for selection into the AllChild programme was shared for the purposes of our matching processes and baseline analysis close to the start of the academic year, once the AllChild participant cohort had been identified.

At baseline, the data provided include:

- AllChild participation status (i.e., our treatment indicator);
- Year group;
- Attendance in the preceding academic year;
- Baseline English (or reading and writing for primary) and maths scores;
- Pupil premium/free school meal (FSM)-eligibility status, Education and Health Care (EHC) plan status, Special Educational Needs (SEN) status, and gender;
- In secondary school, MyVoice survey scores;
- In primary school, teachers' ranking of potential participants' perceived risk of emotional and peer problems (with all those not ranked bottom coded).

Using these data, UCL carried out the matching exercises (as described above) and returned a list of comparator pupils from whom we asked AllChild to collect outcomes data (alongside their routine collection of outcomes data from those participating in their programmes). In addition, baseline SDQ scores were collected for those selected as participant or comparator pupils. We also use these data to characterise participants in the AllChild programme at various points of our analyses.

Data from participants and identified comparator pupils (methods of identifying such pupils set out later in the report) were then shared with UCL again at the end of the academic year (as mid-point outcomes) and the subsequent academic year (as end-point outcomes) to feed into our impact estimation.

Outcome measures collected from children and young people in primary schools (subject to some variation between schools and evaluation years) were:

- % Attendance
- Attainment factor score (providing an average measure of attainment across available measures of reading, writing and maths; discussed further below)
- Strengths and Difficulties Questionnaire (SDQ) scores (total risk score and components)
- Whether or not pupils are considered at risk according to AllChild's risk criteria:
    - Reading (definition differs by school)
    - Writing (definition differs by school)
    - Maths (definition differs by school)

    - Attendance (attendance below 96%)

For secondary school the outcome measures differ slightly (with the same caveat about some variation between schools and over time):

- % Attendance
- Attainment factor score (providing an average measure of attainment across available measures of English and maths; discussed further below)
- SDQ scores (total risk and components)
- MyVoice survey scores
- Confidence survey scores
- Whether or not pupils are considered at risk according to AllChild's risk criteria:
    - English
    - Maths
    - Attendance (attendance below 96%)

An important challenge in the analysis of these data is that the measures of attainment differ between schools participating in the AllChild programme and, in some cases, over time. In order to allow for comparison across schools, we created a single measure of attainment within each school cohort using principal component analysis of the available measures of attainment, extracting the first principal component, and standardising this to have mean zero and standard deviation one. Given that each treated pupil is matched to one (or more) comparison pupils in the same school cohort, the differences between them will be on the basis of the same measures of attainment. However, we do note that bringing together these data across schools and cohorts for pooled analyses does present potential challenges in terms of comparability, which are an important caveat to analyses of these outcomes.

3.4 Statistical description of AllChild cohort

In this section, we use quantitative data to describe the average characteristics of those who are selected to participate in an AllChild early intervention programme, drawing a comparison with the rest of the cohort in participating schools. We do this using the baseline data collected by AllChild — so there can be no potential impact of the programme on the differences that we describe — pooled across all four years of the evaluation.

We report the average characteristics of participants and non-participants in Table 2 for primary school participants and Table 3 for secondary school participants. In both cases, it is very clear that children and young people selected for the AllChild programme are substantially more disadvantaged than their classmates along multiple dimensions.

Table 2: Participant and non-participant characteristics: primary schools

Variable	Background mean	Treated mean	Std. diff	Background sample	Treated sample
Pupil Premium prop.	0.25	0.60	0.796	7676	1395
Special Educational Needs prop.	0.12	0.35	0.644	7718	1397
Male prop.	0.49	0.55	0.120	7718	1397
Pupil wellbeing rank	25.17	11.60	-0.906	7719	1398
Attendance	93.58	91.19	-0.300	7688	1389
Attainment score	0.02	-0.15	-0.114	7719	1398

Notes. Background mean reports average among all pupils in schools where AllChild operates; treated mean reports average among pupils participating in AllChild programme. Std. diff (standardised difference) reports the difference in these two means divided by the pooled standard deviation of the measure.

Table 3: Participant and non-participant characteristics: secondary schools

Variable	Background mean	Treated mean	Std. diff	Background sample	Treated sample
Pupil Premium prop.	0.29	0.63	0.736	4002	394
Special Educational Needs prop.	0.13	0.33	0.580	3995	394
Male prop.	0.61	0.61	0.000	3971	393
Attendance	94.54	92.17	-0.352	4057	395
Attainment score	0.05	-0.20	-0.190	4040	393
SDQ Total Difficulties Score	10.61	16.65	1.007	2959	338

Notes. Background mean reports average among all pupils in schools where AllChild operates; treated mean reports average among pupils participating in AllChild programme. Std. diff (standardised difference) reports the difference in these two means divided by the pooled standard deviation of the measure.

We do not find any evidence of substantial changes in the composition of those participating in the AllChild programme over time, particularly compared to changes in the cohort as a whole.

3.5 Statistical description of matched evaluation sample compared to AllChild and wider cohorts

Given these substantial differences between AllChild participants and their classmates, it would not be a fair test of the programme simply to compare the outcomes of participants

and non-participants and treat the differences as reflecting the effects of participating in the programme. It is for this reason that we, instead, assemble a comparison group who are much more similar to AllChild participants than their classmates as a whole. In this section, we report on this selection process, including the extent to which the matched comparison groups that we have assembled are similar to the relevant group of participants.

We first report on the propensity score estimation process. As discussed above, this proceeds by estimating a model of participation in the AllChild programme using the same data available as part of cohort selection processes. We carried out this modelling (and subsequent matching) on a school-by-school and cohort-by-cohort basis, so here we are not reporting on the propensity score models themselves but rather a proxy using the pooled data for illustrative purposes.

The set of variables included in the propensity score models aimed to include the following covariates, which are based on the measures that are available during AllChild/school cohort selection meetings:

- Year group dummy variables;
- Baseline attendance, squared baseline attendance (to allow for non-linear relationship), and dummy variables interacted with the former two terms to allow for discontinuities at announced risk thresholds;
- Baseline attainment (primary schools: announced risk thresholds, continuous attainment score, and squared attainment score; secondary schools: English scores, squared English scores, English risk thresholds, interactions between English and the English risk thresholds, interactions between English-squared and the English risk thresholds, maths scores, squared maths scores, maths risk thresholds, interactions between maths and the maths risk thresholds, and interactions between maths-squared and the maths risk thresholds);
- Binary variables for pupil premium status, SEN status, and gender;
- SDQ sub-scales (same approach to non-linearity as baseline attendance including with announced risk thresholds for emotional problems scale and peer problems scale) – secondary schools only as these are not collected in primary schools with the full cohort;
- MyVoice survey scores (same approach to non-linearity as baseline attendance including with announced risk thresholds for school engagement risk and parent engagement risk) – secondary schools only as these are not collected in primary schools;
- Ranking of teacher perceived risk of emotional and peer problems (with all those not ranked bottom coded) and squared ranking (to allow for non-linear relationship) – primary only as this is not collected in secondary schools.

Not all of these planned covariates were able to be included in estimating propensity scores in the case of all schools. This was due to a combination of reasons, but primarily

due to lack of variation in continuous variables around announced risk thresholds – for example few children being close to the announcement low-risk threshold for attendance – which would result in collinearity and sample loss if we attempted to retain such variables.

To illustrate the importance of various characteristics as part of this modelling – and sense check the modelling – we present the association between key characteristics and selection into the AllChild programme. We stress that these are not the school-by-school models used in the matching themselves and are simplified for illustrative purposes. The models are reported in Table 4 (for primary) and Table 5 (for secondary).

Table 4: Associations between pupil characteristics and probability of selection onto AllChild programme: primary school

	Change in Selection Prob.
Year 2	0.0560***
	(5.04)
Year 3	0.0557***
	(4.98)
Year 4	0.0445***
	(4.17)
Year 5	0.0443***
	(4.19)
Year 6	-0.0304
	(-1.06)
Attendance %	-0.00385***
	(-6.87)
Attainment	-0.00367
	(-1.57)
Pupil premium	0.131***
	(20.31)
Male	0.0139*
	(2.06)
SEN	0.0852***
	(11.01)
Wellbeing Ranking	-0.00701***
	(-24.18)
N	9032

Notes. Reporting average marginal effects from underlying probit regression model; t statistics in parentheses; * p < 0.05, ** p < 0.01, *** p < 0.001.

Table 5: Associations between pupil characteristics and probability of selection onto AllChild programme: secondary school

	Change in Selection Prob.
Year 8	-0.0250
	(-1.89)
Year 9	-0.0611***
	(-4.98)
Attendance %	-0.00540***
	(-4.91)
Attainment	-0.00779*
	(-2.10)
Pupil premium	0.0857***
	(8.51)
Male	0.0161
	(1.53)
SEN	0.0602***
	(5.33)
Confidence Score	-0.00978
	(-0.91)
SDQ Total Score	0.00754***
	(7.58)
MyVoice Wellbeing Score	0.00797
	(0.94)
MyVoice School Engagement Score	0.0405***
	(5.20)
MyVoice Parent Engagement Score	-0.0359**
	(-3.28)
MyVoice Social Support Score	0.0106
	(1.41)
MyVoice Anti-Social Score	-0.000687
	(-0.10)

N	2856

Notes. Reporting average marginal effects from underlying probit regression model; t statistics in parentheses; * p < 0.05, ** p < 0.01, *** p < 0.001.

The modelling indicates that the following groups are more likely to be selected for participation in the AllChild programme, which fits with our understanding of its aims and targeting: boys (in primary), pupils eligible for free school meals (an indicator of coming from a low income family), pupils with special educational needs, those with worse attendance (although in our full modelling we include a more flexible relationship capturing that selection probability declines again for worse attendance) and those with potential wellbeing issues (indicated by, for example, wellbeing ranking for primary and SDQ for secondary). As noted, this is an indicative model rather than that used in the matching process itself, but it does suggest our propensity scores are capturing important predictors of participation that are also known to be predictive of pupils' other expected outcomes.

Turning next to the outcome of the matching process, we do similarly to how we reported differences between participants and non-participants in the previous section, but also provide what is known as a 'standardised difference' (or, in the context of impact estimates, an 'effect size') between participants and the matched comparator groups. We are looking to achieve as small standardised differences as possible, with a common rule of thumb being that standardised differences of less than 0.1 are fairly well matched.

We report this remaining imbalance separately for our primary and secondary school samples, and for the 'wide' and 'narrow' matched comparison groups for each as described earlier in the report.

Table 6: Matched participant and non-participant characteristics: primary school wide-matched sample

Variable	Background mean	Comparison mean	Treated mean	Std. diff	Comparison sample	Treated sample
Pupil Premium prop.	0.25	0.42	0.60	0.367	2145	1349
Special Educational Needs prop.	0.12	0.20	0.35	0.347	2145	1349
Male prop.	0.49	0.51	0.55	0.080	2145	1349
Pupil wellbeing rank	25.17	17.04	11.64	-0.467	2145	1349
Attendance	93.58	92.65	91.27	-0.183	2145	1349
Attainment score	0.02	-0.13	-0.14	-0.006	2145	1349
SDQ Total Difficulties Score	N/A	8.73	12.49	0.563	1831	1290

Notes. Background mean reports average among all pupils in schools where AllChild operates, except participants; comparison mean reports average among pupils identified as matched comparators for pupils participating in AllChild programme;

treated mean reports average among pupils participating in AllChild programme. Std. diff (standardised difference) reports the difference between comparison and treated means divided by the pooled standard deviation of the measure; prop. = proportion.

Table 7: Matched participant and non-participant characteristics: primary school narrow-matched sample

Variable	Background mean	Comparison mean	Treated mean	Std. diff	Comparison sample	Treated sample
Pupil Premium prop.	0.25	0.46	0.55	0.180	806	706
Special Educational Needs prop.	0.12	0.27	0.28	0.022	806	706
Male prop.	0.49	0.53	0.51	-0.040	806	706
Pupil wellbeing rank	25.17	15.97	14.34	-0.137	806	706
Attendance	93.58	92.34	91.73	-0.077	806	706
Attainment score	0.02	-0.17	-0.19	-0.012	806	706
SDQ Total Difficulties Score	N/A	9.60	11.29	0.255	679	675

Notes. Background mean reports average among all pupils in schools where AllChild operates, except participants; comparison mean reports average among pupils identified as matched comparators for pupils participating in AllChild programme; treated mean reports average among pupils participating in AllChild programme. Std. diff (standardised difference) reports the difference between comparison and treated means divided by the pooled standard deviation of the measure; prop. = proportion.

Table 8: Matched participant and non-participant characteristics: secondary school wide-matched sample

Variable	Background mean	Comparison mean	Treated mean	Std. diff	Comparison sample	Treated sample
Pupil Premium prop.	0.29	0.49	0.63	0.284	778	389
Special Educational Needs prop.	0.13	0.22	0.33	0.255	778	389
Male prop.	0.61	0.63	0.60	-0.062	778	389
Attendance	94.54	93.67	92.17	-0.197	778	389
Attainment score	0.05	-0.11	-0.20	-0.066	768	387
SDQ Total Difficulties Score	10.61	13.60	16.66	0.479	625	336

Notes. Background mean reports average among all pupils in schools where AllChild operates, except participants; comparison mean reports average among pupils identified as matched comparators for pupils participating in AllChild programme;

treated mean reports average among pupils participating in AllChild programme. Std. diff (standardised difference) reports the difference between comparison and treated means divided by the pooled standard deviation of the measure; prop. = proportion.

Table 9: Matched participant and non-participant characteristics: secondary school narrow-matched sample

Variable	Background mean	Comparison mean	Treated mean	Std. diff	Comparison sample	Treated sample
Pupil Premium prop.	0.29	0.56	0.63	0.143	285	242
Special Educational Needs prop.	0.13	0.26	0.29	0.067	285	242
Male prop.	0.61	0.67	0.61	-0.125	285	242
Attendance	94.54	92.99	92.00	-0.142	285	242
Attainment score	0.05	-0.12	-0.36	-0.170	281	240
SDQ Total Difficulties Score	10.61	14.83	15.71	0.135	223	199

Notes. Background mean reports average among all pupils in schools where AllChild operates, except participants; comparison mean reports average among pupils identified as matched comparators for pupils participating in AllChild programme; treated mean reports average among pupils participating in AllChild programme. Std. diff (standardised difference) reports the difference between comparison and treated means divided by the pooled standard deviation of the measure; prop. = proportion.

These tables demonstrate that, while we have successfully been able to identify comparison groups that are a substantial improvement on attempting to evaluate the impact of AllChild simply by comparing participants with non-participants, we caution that observable (and, likely unobservable) differences do remain between participants in the treatment and matched comparison groups that we will be using for our analyses. This is the case even where we have taken quite an aggressive approach to identifying well-matched comparison groups at the expense of sample size – ultimately young people selected into AllChild are extremely different to their peer group and, as a result, it is extremely difficult to identify a suitable number of truly comparable individuals as a comparison group. The bias is, however, particularly pronounced in the 'wide' matched sample and, as such, we would advise that results based on the 'narrow' matched sample are more likely to be a reliable guide when available.

Using these matched samples, we are unlikely to be able to recover impact estimates that impartial observers would consider to be truly causal, but could still provide evidence that we judge likely to be persuasive. This is particularly the case because the comparison groups assembled generally have average characteristics that are correlated with stronger

outcomes (educational and wider) than the AllChild participants to whom they will be compared. Hence, the directional bias we might expect on our estimates of treatment would be for an underestimate, not an overestimate, of treatment. Any negative estimates should be interpreted with this in mind, while we judge that positive estimates are unlikely to be undermined by selection bias.

3.6 Statistical description of discontinuity evaluation sample compared to AllChild and wider cohorts

We also examine the balance among the sample assembled for our discontinuity-based evaluation design. As noted above, unlike the matching design, this work focuses only on primary schools as initial exploration found that the proposed discontinuity did not effect the required change in probability of selection in secondary school participants.

First, we look at young people either side of the 96% attendance threshold that is used as part of the AllChild cohort selection process. Consistent with our matching approach, we defined two windows around this threshold to explore potential trade-offs of sample size and imbalance.

After exploring various alternatives, we defined the narrow sample as those between 95.5% and 96.5% attendance (i.e., a window of 1% with 0.5% above and 0.5% below the threshold) and the wide sample as those between 95% and 97% attendance (i.e., a window of 2% with 1% above and 1% below the threshold). For primary schools, the narrow sample contains 590 individuals, while the wide sample contains 805 individuals.

Table 10: Numbers of participants and non-participants just either side of attendance threshold: narrow sample

	AllChild	Participant	
Above Threshold	No	Yes	Total
No	258	36	294
Yes	240	56	296
Total	498	92	590

Table 11: Numbers of participants and non-participants just either side of attendance threshold: wide sample

	AllChild	Participant	
Above Threshold	No	Yes	Total
No	355	47	402
Yes	322	81	403

Total	677	128	805

As can be seen from Table 10 (for narrow sample) and Table 11 (for wide sample), there are roughly the same number of young people overall above and below the threshold within each window, suggesting that there is no particular bunching of this measure either side of the measure (which would suggest it was being manipulated in some way).

Our first concern is whether there is evidence that being just below, rather than just above, the 96% attendance threshold increases the probability of selection onto the AllChild programme. Without this jump in probability at this point there not be further meaningful analysis that we can do with this method. This is also evident by looking at Table 12 and Table 13, looking at the numbers above and below the threshold by whether or not they are AllChild participants. However, we can see this more intuitively by looking at these as percentages.

Table 12: Numbers of participants and non-participants just either side of attendance threshold: narrow sample

	AllChild	Participant	
Above Threshold	No	Yes	Total
No	88%	12%	100%
Yes	81%	19%	100%
Total	84%	16%	100%

Table 13: Numbers of participants and non-participants just either side of attendance threshold: wide sample

	AllChild	Participant	
Above Threshold	No	Yes	Total
No	88%	12%	100%
Yes	80%	20%	100%
Total	84%	16%	100%

We can see in Table 12 (for narrow sample) and Table 13 (for wide sample) that those who fall just below the 96% threshold are (for either sample) 7-8%pts more likely to be selected onto the AllChild programme than otherwise similar young people with just slightly higher attendance. It is this discontinuous change in selection probability that we will be using to compare differences in young people's outcomes.

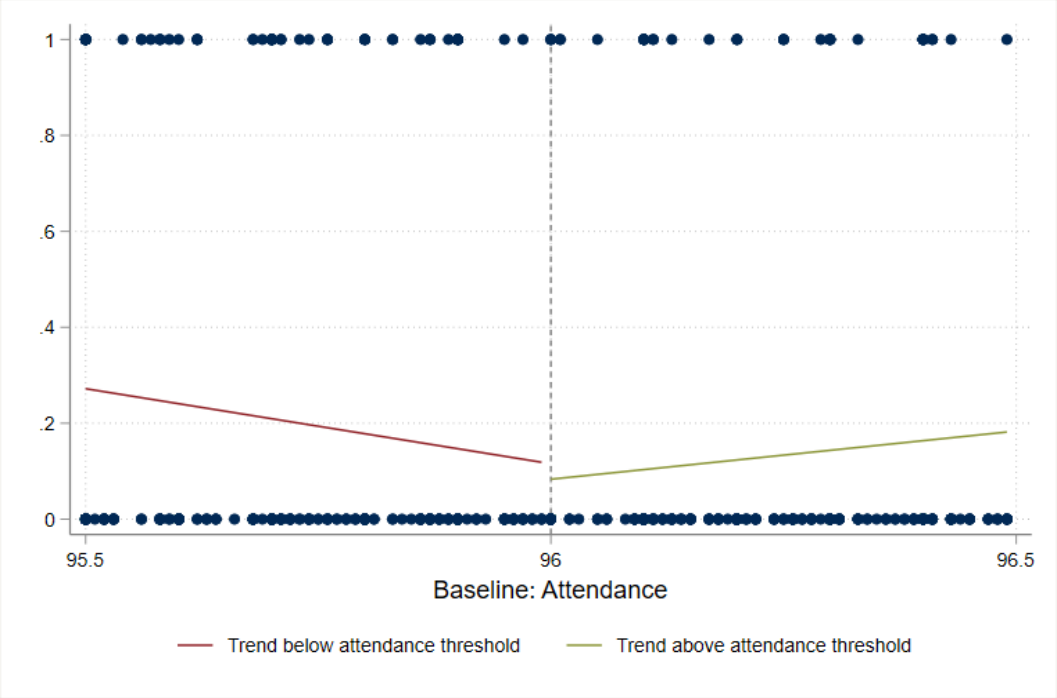

Figure 1: Probability of selection onto AllChild programme by attendance around 96% threshold — narrow sample

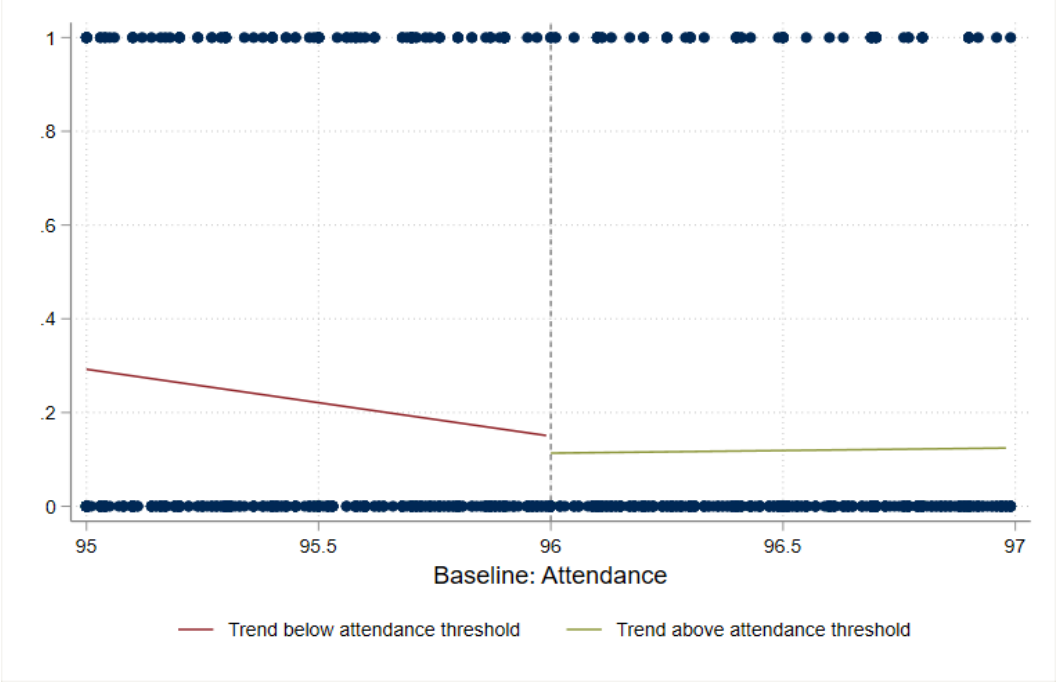

Figure 2: Probability of selection onto AllChild programme by attendance around 96% threshold — wide sample

We can also look at this graphically (Figure 1 for narrow sample; Figure 2 for wide sample), where we go slightly further in using a separate linear fit either side of the threshold to estimate the jump at the exact point of the threshold (rather than averaged across the width of the sample selected either side of the threshold). Here we can see this same jump down in probability of selection at the point of the 96% attendance threshold.

Is there evidence that those on either side of this threshold are similar in other ways, too? We consider this by looking at imbalance in a way that will be familiar from the matched sample analysis.

Table 14: Participant and non-participant characteristics around 96% discontinuity: narrow sample

Variable	Background mean	Below mean	Above mean	Std. diff	Below sample	Above sample
Pupil Premium prop.	0.25	0.25	0.31	-0.135	297	295
Special Educational Needs prop.	0.12	0.18	0.18	0.000	297	295
Male prop.	0.49	0.51	0.51	0.000	297	295
Pupil wellbeing rank	25.17	21.97	21.13	0.056	297	295
Attainment score	0.02	-0.05	0.13	-0.125	297	295
SDQ Total Difficulties Score	N/A	8.32	8.14	0.027	257	263
Propensity score	0.15	0.17	0.13	0.156	297	295
Attendance	93.58	95.76	96.19	-3.071	297	295
Selected for AllChild cohort	0.00	0.19	0.12	0.194	296	294

Notes. Background mean reports average among all pupils in schools where AllChild operates, except participants; below mean reports average among pupils just below 96% attendance threshold; above mean reports average among pupils just above 96% attendance threshold. Std. diff (standardised difference) reports the difference between below and above means divided by the pooled standard deviation of the measure; prop. = proportion.

Table 15: Participant and non-participant characteristics around 96% discontinuity: wide sample

Variable	Background mean	Below mean	Above mean	Std. diff	Below sample	Above sample
Pupil Premium prop.	0.25	0.30	0.31	-0.022	404	403

Special Educational Needs prop.	0.12	0.18	0.17	0.027	404	403
Male prop.	0.49	0.50	0.52	-0.040	404	403
Pupil wellbeing rank	25.17	21.33	19.80	0.103	404	403
Attainment score	0.02	-0.07	0.03	-0.069	404	403
SDQ Total Difficulties Score	N/A	8.39	8.01	0.057	354	359
Propensity score	0.15	0.18	0.14	0.153	404	403
Attendance	93.58	95.64	96.32	-2.720	404	403
Selected for AllChild cohort	0.00	0.20	0.12	0.221	403	402

Notes. Background mean reports average among all pupils in schools where AllChild operates, except participants; below mean reports average among pupils just below 96% attendance threshold; above mean reports average among pupils just above 96% attendance threshold. Std. diff (standardised difference) reports the difference between below and above means divided by the pooled standard deviation of the measure; prop. = proportion.

Compared to the matched analysis samples, it is notable that, as we would hope for this kind of analysis, there is much less evidence of imbalance in young people's characteristics either side of the attendance threshold. This is not notably more the case in the narrow sample (Table 14) than the wide sample (Table 15). The notable exceptions for this in the tables are for the characteristics we would expect: attendance is different because our threshold is based on attendance, and selection for AllChild cohort is different, which is what we are hoping to be the case for our design to work and as we documented above.

4 Overall results

In this section, we report estimated impacts of the AllChild programme based on our two impact evaluation methods. Here, we will focus on overall (average) impacts, while in the [discussion][subsequent section] we will explore whether there is evidence that our estimates vary based on aspects of programme delivery.

4.1 Overall differences between AllChild and matched comparison cohorts

We begin with results from the matched design, reporting analyses bringing together data from all evaluation cohorts. In each case, we report on the availability of outcomes data for out treated and matched comparison samples (and, equivalently, the attrition from our analysis sample compared to those selected to be part of the evaluation analysis based on the matching process at baseline), before proceeding to report differences in outcome measures (first at mid-point, then at end-point) between the treated and matched comparison groups. As described above, we have two such groups: one a 'wide' matched sample (with a larger sample size, but more issues with imbalance, as set out earlier) and the other a 'narrow' matched sample (with a smaller sample size, but better balance between the treatment and comparison groups).

The results themselves are reported in tables in which each column corresponds to an outcome measure of interest and the rows report:

- The raw treatment estimate (i.e., the expected change in the outcome measure among members of the AllChild cohort compared to matched comparison group members)
- The p-value for this treatment estimate in parentheses (i.e., a measure of the statistical significance of the treatment estimate where lower means less likely due to chance and figures below 0.05 are typically considered 'statistically significant')
- N (i.e., the sample size for the analysis in question, which will correspond to the 'achieved sample' in the attrition tables)
- Cohen's d (i.e., a widely used standardised effect size measure, which places treatment estimates for different outcome measures on the same scale)
- And, where relevant (i.e., for binary risk outcomes) the percentage change in the probability of the outcome implied by the treatment estimate.

As discussed in Section 3, we do caution that results must be interpreted in light of the residual confounding present in the sample that is likely to bias effects in a negative direction, especially in the wide matched sample. In this section we will report the results that are estimated but will contextualise these with this in mind in drawing our conclusions.

4.1.1 Pooled analysis of all available mid-point outcomes

Considering mid-point outcomes, we bring together data from all four evaluation cohorts where possible.

4.1.1.1 Primary schools

Across participating primary schools and most of our outcome measures we achieve a sample of around 2,500 in the wide matched sample and just over 1,000 in the narrow matched sample (attrition of around a quarter of the matched samples). This is reduced further in the case of SDQ due to non-participation of some schools in these outcome measures at this time point, with attrition rising above 50%.

Table 16: Attrition from mid-point outcome measures in primary schools

Wide Match			
Outcome	Expected sample	Achieved sample	Attrition %
Attendance At Risk	3494	2596	25.70
Writing At Risk	3494	2534	27.48
Reading At Risk	3494	2509	28.19
Maths At Risk	3494	2541	27.28
Attainment Score	3494	2564	26.62
Attendance	3494	2596	25.70
SDQ	3494	1446	58.61
SDQ: Emotional Problems	3494	1453	58.41
SDQ: Conduct Problems	3494	1452	58.44
SDQ: Hyperactivity	3494	1453	58.41
SDQ: Peer Problems	3494	1453	58.41
SDQ: Pro-social	3494	1453	58.41
SDQ: Internalising	3494	1453	58.41
SDQ: Externalising	3494	1452	58.44
Narrow Match			
Outcome	Expected sample	Achieved sample	Attrition %
Attendance At Risk	1512	1108	26.72
Writing At Risk	1512	1075	28.90
Reading At Risk	1512	1068	29.37
Maths At Risk	1512	1075	28.90
Attainment Score	1512	1092	27.78
Attendance	1512	1108	26.72
SDQ	1512	702	53.57
SDQ: Emotional Problems	1512	705	53.37
SDQ: Conduct Problems	1512	705	53.37
SDQ: Hyperactivity	1512	705	53.37

SDQ: Peer Problems	1512	705	53.37
SDQ: Pro-social	1512	705	53.37
SDQ: Internalising	1512	705	53.37
SDQ: Externalising	1512	705	53.37

Among the available sample, we report the estimated difference in outcome measures associated with participating in the AllChild programme compared to the matched sample in Table 17 (wide matched sample) and Table 18 (narrow matched sample).

In the case of attendance and SDQ outcomes, our findings are consistently in the direction for which we would hope: positive for attendance and negative for SDQ (where higher scores indicate potential cause for concern). However, the effects are small (0.04-0.06) and are not statistically significant across either outcome or matched sample.

The picture with attainment is more mixed: the estimate is positive in the case of the wide matched sample but negative in the narrow matched sample (contrary to expectations in terms of direction of any selection bias on performance). However, the results are extremely small (effect size of 0.02-0.03) and not statistically significant, so it is probable that the difference between them is caused by random variation.

Table 17: Impacts of AllChild participation at mid-point: wide matched sample

	Attendance	Attainment	SDQ
Cohort Membership	0.300	0.0364	-0.417
	(0.36)	(0.48)	(0.10)
N	2596	2564	1446
Cohen's d	0.0400	0.0200	-0.0600

Notes. Estimated differences in outcome measures associated with cohort membership obtained from regression models; p-values in parentheses adjusted for evaluation design clustering.

Table 18: Impacts of AllChild participation at mid-point: narrow matched sample

	Attendance	Attainment	SDQ
Cohort Membership	0.473	-0.0451	-0.430
	(0.26)	(0.51)	(0.22)
N	1108	1092	702
Cohen's d	0.0600	-0.0300	-0.0600

Notes. Estimated differences in outcome measures associated with cohort membership obtained from regression models; p-values in parentheses adjusted for evaluation design clustering.

Feeding into that mixed picture on attainment, we estimate statistically significant increases in participants being judged as at risk across reading, writing and maths in the wide matched sample. These findings only persist for writing risk in the narrow matched sample (where we would expect less bias in our findings), but this does point to a potential concern worth considering.

Table 19: Impacts of AllChild participation at mid-point: wide matched sample

	Read Risk	Write Risk	Maths Risk	Attend. Risk
Cohort Membership	0.133	0.150	0.161	0.0198
	(0.00)	(0.00)	(0.00)	(0.31)
N	2509	2534	2541	2596
Cohen's d	0.280	0.310	0.340	0.0400
% Change	13.35	15.04	16.07	1.980

Notes. Estimated differences in outcome measures associated with cohort membership obtained from regression models; p-values in parentheses adjusted for evaluation design clustering.

Table 20: Impacts of AllChild participation at mid-point: narrow matched sample

	Read Risk	Write Risk	Maths Risk	Attend. Risk
Cohort Membership	0.0325	0.0614	0.0473	0.0427
	(0.25)	(0.02)	(0.11)	(0.15)
N	1068	1075	1075	1108
Cohen's d	0.0700	0.130	0.100	0.0900
% Change	3.250	6.140	4.730	4.270

Notes. Estimated differences in outcome measures associated with cohort membership obtained from regression models; p-values in parentheses adjusted for evaluation design clustering.

By contrast, findings are considerably more encouraging when it comes to scores in the Strengths and Difficulties Questionnaire (SDQ). In both the wide and narrow matched samples, the results point to statistically significant reductions in the Peer Problems sub-scale and increases in the ProSocial scale. Similar estimates have been present fairly

consistently in year-by-year analysis of midpoint outcomes across all but 2021 (when this measure was not analysed at midpoint as it was not generally collected). As noted above, these are not quite enough to shift the dial in terms of the overall SDQ scores, but point to potentially important improvements in this domain.

Table 21: Impacts of AllChild participation at mid-point: wide sample

	Emotional Prob.	Conduct Prob.	Hyper-activity	Peer Probs	Pro Social	Internal-ising	External-ising	Total
Cohort Membership	0.0833	-0.155	0.0615	-0.396	0.448	-0.312	-0.0940	-0.417
	(0.51)	(0.13)	(0.65)	(0.00)	(0.00)	(0.05)	(0.64)	(0.10)
N	1453	1452	1453	1453	1453	1453	1452	1446
Cohen's d	0.0400	-0.0800	0.0200	-0.230	0.180	-0.0900	-0.0200	-0.0600

Notes. Estimated differences in outcome measures associated with cohort membership obtained from regression models; p-values in parentheses adjusted for evaluation design clustering.

Table 22: Impacts of AllChild participation at mid-point: narrow matched sample

	Emotional Prob.	Conduct Prob.	Hyper-activity	Peer Probs	Pro Social	Internal-ising	External-ising	Total
Cohort Membership	0.128	-0.120	-0.00373	-0.437	0.493	-0.309	-0.123	-0.430
	(0.47)	(0.39)	(0.98)	(0.00)	(0.00)	(0.19)	(0.67)	(0.22)
N	705	705	705	705	705	705	705	702
Cohen's d	0.0500	-0.0600	0	-0.250	0.200	-0.0900	-0.0300	-0.0600

Notes. Estimated differences in outcome measures associated with cohort membership obtained from regression models; p-values in parentheses adjusted for evaluation design clustering.

4.1.1.2 Secondary schools

Turning to mid-point outcomes in secondary schools, across most of our outcome measures we achieve a sample of around 600-800 in the wide matched sample and 200-300 in the narrow matched sample (attrition of between a third and a half) in the case of core outcomes of attendance and attainment. This reflects a smaller number of participating secondary schools, only partially offset by larger cohort sizes, including no secondary schools in the fourth and final evaluation cohort.

Table 23: Attrition from mid-point outcome measures in secondary schools

Wide Match			
Outcome	Expected sample	Achieved sample	Attrition %
Attendance	1167	832	28.71
Attainment Score	1167	815	30.16
English At Risk	1167	649	44.39
Maths At Risk	1167	666	42.93
Attendance At Risk	1167	832	28.71
Narrow Match			
Outcome	Expected sample	Achieved sample	Attrition %
Attendance	527	367	30.36
Attainment Score	527	334	36.62
English At Risk	527	267	49.34
Maths At Risk	527	273	48.20
Attendance At Risk	527	367	30.36

The results themselves are reported in Table 24 for the wide matched sample and Table 25 for the narrow matched sample. Results are less encouraging for participants in secondary schools. Attendance is positive (but insignificant) in the wide matched sample but tips across to negative (but very small and insignificant) in the narrow matched sample. A statistically significant negative impact is estimated on attainment in the wide sample, albeit that this loses statistical significance (but not much of its magnitude) in the narrow matched sample. Estimated effects on risk thresholds are consistent with the overall picture for attainment.

Table 24: Impacts of AllChild participation at mid-point: wide matched sample

	Attendance	Attainment	English Risk	Maths Risk	Attendance Risk
Cohort Membership	0.629	-0.194	0.0676	0.113	0.0192
	(0.46)	(0.01)	(0.07)	(0.01)	(0.56)
N	832	815	649	666	832
Cohen's d	0.0500	-0.160	0.140	0.230	0.0400
% Change			6.760	11.32	1.920

Notes. Estimated differences in outcome measures associated with cohort membership obtained from regression models; p-values in parentheses adjusted for evaluation design clustering.

Table 25: Impacts of AllChild participation at mid-point: narrow matched sample

	Attendance	Attainment	English Risk	Maths Risk	Attendance Risk
Cohort Membership	-0.107	-0.173	0.0633	0.0799	0.0291
	(0.92)	(0.12)	(0.26)	(0.19)	(0.48)
N	367	334	267	273	367
Cohen's d	-0.0100	-0.140	0.130	0.160	0.0600
% Change			6.330	7.990	2.910

Notes. Estimated differences in outcome measures associated with cohort membership obtained from regression models; p-values in parentheses adjusted for evaluation design clustering.

4.1.2 Pooled analysis of all available end-point outcomes

Turning next to end-point outcomes, we use data from the first three evaluation cohorts (the fourth cohort have not reached end-point measurement at the time of this report's analysis).

4.1.2.1 Primary schools

In primary school endpoint outcomes, across our outcome measures we achieve a sample of around 1,600-1,700 pupils in the wide matched sample and around 650 in the narrow matched sample (attrition of around 40% of the matched samples). It is worth noting that this is substantially more attrition than was the case for mid-point outcomes, which is important context for differences between findings at these two time points. Conversely, however, for SDQ outcomes the attrition is actually slightly less marked than at mid-point, due to cohorts in which SDQ was not collected at mid-point but was collected at end-point.

Table 26: Attrition from end-point outcome measures in primary schools

Wide Match			
Outcome	Expected sample	Achieved sample	Attrition %
Attendance At Risk	2803	1704	39.21
Writing At Risk	2803	1579	43.67
Reading At Risk	2803	1585	43.45
Maths At Risk	2803	1566	44.13
Attainment Score	2803	1619	42.24
Attendance	2803	1704	39.21
SDQ	2803	1703	39.24

SDQ: Emotional Problems	2803	1703	39.24
SDQ: Conduct Problems	2803	1703	39.24
SDQ: Hyperactivity	2803	1703	39.24
SDQ: Peer Problems	2803	1703	39.24
SDQ: Pro-social	2803	1703	39.24
SDQ: Internalising	2803	1703	39.24
SDQ: Externalising	2803	1703	39.24
Narrow Match			
Outcome	Expected sample	Achieved sample	Attrition %
Attendance At Risk	1079	650	39.76
Writing At Risk	1079	605	43.93
Reading At Risk	1079	616	42.91
Maths At Risk	1079	613	43.19
Attainment Score	1079	634	41.24
Attendance	1079	650	39.76
SDQ	1079	663	38.55
SDQ: Emotional Problems	1079	663	38.55
SDQ: Conduct Problems	1079	663	38.55
SDQ: Hyperactivity	1079	663	38.55
SDQ: Peer Problems	1079	663	38.55
SDQ: Pro-social	1079	663	38.55
SDQ: Internalising	1079	663	38.55
SDQ: Externalising	1079	663	38.55

Then comparing performance between AllChild participants and non-participants, we find only extremely small differences when using the wide matched sample. Some small differences in a negative direction for attainment and SDQ do emerge in the narrow sample, but they are still very small and far from statistical significance. Practically there is no evidence of difference at all across primary school participants at the endpoint.

Table 27: Impacts of AllChild participation at end-point: wide matched sample

	Attendance	Attainment	SDQ
Cohort Membership	0.156	-0.0157	0.151
	(0.61)	(0.81)	(0.61)
N	1704	1619	1703
Cohen's d	0.0200	-0.0100	0.0200

Notes. Estimated differences in outcome measures associated with cohort membership obtained from regression models; p-values in parentheses adjusted for evaluation design clustering.

Table 28: Impacts of AllChild participation at end-point: narrow matched sample

	Attendance	Attainment	SDQ
Cohort Membership	0.0646	-0.0839	0.317
	(0.86)	(0.32)	(0.48)
N	650	634	663
Cohen's d	0.0100	-0.0500	0.0500

Notes. Estimated differences in outcome measures associated with cohort membership obtained from regression models; p-values in parentheses adjusted for evaluation design clustering.

The findings for attainment risk are broadly consistent with those found at mid-point, with statistically significant negative findings in the wide matched sample being reduced to insignificance in the narrow matched sample.

Table 29: Impacts of AllChild participation at end-point: wide matched sample

	Read Risk	Write Risk	Maths Risk	Attend. Risk
Cohort Membership	0.0777	0.0952	0.0574	0.0305
	(0.00)	(0.00)	(0.01)	(0.19)
N	1585	1579	1566	1704
Cohen's d	0.200	0.220	0.150	0.0600
% Change	7.770	9.520	5.740	3.050

Notes. Estimated differences in outcome measures associated with cohort membership obtained from regression models; p-values in parentheses adjusted for evaluation design clustering.

Table 30: Impacts of AllChild participation at end-point: narrow matched sample

	Read Risk	Write Risk	Maths Risk	Attend. Risk
Cohort Membership	0.0330	0.0109	-0.00187	0.0286

	(0.31)	(0.74)	(0.95)	(0.41)
N	616	605	613	650
Cohen's d	0.0800	0.0300	0	0.0600
% Change	3.300	1.090	-0.190	2.860

Notes. Estimated differences in outcome measures associated with cohort membership obtained from regression models; p-values in parentheses adjusted for evaluation design clustering.

Unfortunately, the positive picture from primary school mid-point in SDQ performance does not translate across to the end-point outcome measures, indeed there is a more concerning statistically significant estimate for emotional problems in the narrow-matched sample at this time point. It is possible that these differencees between mid-point and end-point are driven by the smaller attrition for SDQ measures at end-point, although this is necessarily speculative.

Table 31: Impacts of AllChild participation at end-point: wide matched sample

	Emotional Prob.	Conduct Prob.	Hyper-activity	Peer Probs	Pro Social	Internal-ising	External-ising	Total
Cohort Membership	0.0745	-0.0229	0.180	-0.0803	0.173	-0.00582	0.157	0.151
	(0.51)	(0.82)	(0.20)	(0.36)	(0.14)	(0.97)	(0.45)	(0.61)
N	1703	1703	1703	1703	1703	1703	1703	1703
Cohen's d	0.0300	-0.0100	0.0600	-0.0500	0.0700	0	0.0400	0.0200

Notes. Estimated differences in outcome measures associated with cohort membership obtained from regression models; p-values in parentheses adjusted for evaluation design clustering.

Table 32: Impacts of AllChild participation at end-point: narrow matched sample

	Emotional Prob.	Conduct Prob.	Hyper-activity	Peer Probs	Pro Social	Internal-ising	External-ising	Total
Cohort Membership	0.381	-0.0619	0.0659	-0.0675	0.161	0.314	0.00396	0.317
	(0.03)	(0.68)	(0.75)	(0.60)	(0.36)	(0.21)	(0.99)	(0.48)
N	663	663	663	663	663	663	663	663
Cohen's d	0.170	-0.0300	0.0200	-0.0400	0.0700	0.0900	0	0.0500

Notes. Estimated differences in outcome measures associated with cohort membership obtained from regression models; p-values in parentheses adjusted for evaluation design clustering.

4.1.2.2 Secondary schools

Turning, finally, to secondary schools at endpoint, we achieve rather variable samples across different outcome measures reflecting differential participation in these by different secondary schools. These vary from attendance (almost 1,000 for the wide matched sample; almost 400 for the narrow matched sample) down to MyVoice survey measures (around 400 for the wide match sample; below 150 for the narrow matched sample). This does, however, allow us to look at these wider outcomes for secondary school participants in way that was not possible at midpoint.

Table 33: Attrition from end-point outcome measures in secondary schools

Wide Match			
Outcome	Expected sample	Achieved sample	Attrition %
Attendance	1563	974	37.68
Attainment Score	1563	752	51.89
English At Risk	1563	677	56.69
Maths At Risk	1563	662	57.65
Attendance At Risk	1563	974	37.68
Confidence Score	1563	800	48.82
Confidence At Risk	1563	806	48.43
MyVoice: Wellbeing Risk	1563	380	75.69
MyVoice: School Engagement Risk	1563	410	73.77
MyVoice: Parent Engagement Risk	1563	369	76.39
MyVoice: Social Support Risk	1563	382	75.56
MyVoice: Anti-Social Risk	1563	383	75.5
SDQ	1563	872	44.21
SDQ: Emotional Problems	1563	872	44.21
SDQ: Conduct Problems	1563	872	44.21
SDQ: Hyperactivity	1563	872	44.21
SDQ: Peer Problems	1563	872	44.21
SDQ: Pro-social	1563	874	44.08
SDQ: Internalising	1563	872	44.21
SDQ: Externalising	1563	872	44.21

Narrow Match Outcome	Expected sample	Achieved sample	Attrition %
Attendance	636	382	39.94
Attainment Score	636	299	52.99
English At Risk	636	269	57.7
Maths At Risk	636	262	58.81
Attendance At Risk	636	382	39.94
Confidence Score	636	284	55.35
Confidence At Risk	636	284	55.35
MyVoice: Wellbeing Risk	636	141	77.83
MyVoice: School Engagement Risk	636	149	76.57
MyVoice: Parent Engagement Risk	636	130	79.56
MyVoice: Social Support Risk	636	138	78.3
MyVoice: Anti-Social Risk	636	136	78.62
SDQ	636	315	50.47
SDQ: Emotional Problems	636	315	50.47
SDQ: Conduct Problems	636	315	50.47
SDQ: Hyperactivity	636	315	50.47
SDQ: Peer Problems	636	315	50.47
SDQ: Pro-social	636	315	50.47
SDQ: Internalising	636	315	50.47
SDQ: Externalising	636	315	50.47

The results themselves are reported in Table 34 for the wide matched sample and Table 35 for the narrow matched sample. Results are somewhat consistent with those seen at midpoint, albeit negative (but insignificant) for attendance across both the wide and narrow matched sample. A statistically significant negative impact is estimated on attainment in the wide sample, albeit that this loses statistical significance (but not much of its magnitude) in the narrow matched sample. An estimated negative impact on English risk persists across wide and narrow matched samples.

Table 34: Impacts of AllChild participation at end-point: wide matched sample

	Attendance	Attainment	English Risk	Maths Risk	Attendance Risk
Cohort Membership	0.781	-0.300	0.0945	0.0728	-0.0221
	(0.29)	(0.00)	(0.01)	(0.03)	(0.42)
N	974	752	677	662	974

Cohen's d	0.0700	-0.200	0.220	0.160	-0.0500
% Change			9.450	7.280	-2.210

Notes. Estimated differences in outcome measures associated with cohort membership obtained from regression models; p-values in parentheses adjusted for evaluation design clustering.

Table 35: Impacts of AllChild participation at end-point: narrow matched sample

	Attendance	Attainment	English Risk	Maths Risk	Attendance Risk
Cohort Membership	-1.790	-0.228	0.109	0.0452	-0.0157
	(0.12)	(0.07)	(0.03)	(0.36)	(0.68)
N	382	299	269	262	382
Cohen's d	-0.160	-0.150	0.250	0.100	-0.0300
% Change			10.90	4.520	-1.570

Notes. Estimated differences in outcome measures associated with cohort membership obtained from regression models; p-values in parentheses adjusted for evaluation design clustering.

We see no evidence of statistically significant impacts across domains of the SDQ (Table 36 and Table 37), nor across most domains of the MyVoice survey (Table 38 and Table 39). There is a concerning statistically significant impact on the AntiSocial domain of MyVoice, but only in the wide matched sample; conversely there is an encouraging statistically significant impact on the Social Support domain, but only in the narrow matched sample (albeit that this is where we would expect less bias).

Table 36: Impacts of AllChild participation at end-point: wide matched sample

	Emotional Prob.	Conduct Prob.	Hyper-activity	Peer Probs	Pro Social	Internal-ising	External-ising	Total
Cohort Membership	-0.0819	0.169	0.180	0.0695	-0.0108	-0.0125	0.349	0.337
	(0.61)	(0.26)	(0.32)	(0.56)	(0.94)	(0.96)	(0.20)	(0.40)
N	872	872	872	872	874	872	872	872
Cohen's d	-0.0300	0.0700	0.0600	0.0400	0	0	0.0800	0.0500

Notes. Estimated differences in outcome measures associated with cohort membership obtained from regression models; p-values in parentheses adjusted for evaluation design clustering.

Table 37: Impacts of AllChild participation at end-point: narrow matched sample

	Emotional Prob.	Conduct Prob.	Hyper-activity	Peer Probs	Pro Social	Internal-ising	External-ising	Total
Cohort Membership	0.107	-0.154	-0.0257	0.0442	0.118	0.151	-0.179	-0.0285
	(0.70)	(0.54)	(0.93)	(0.85)	(0.64)	(0.71)	(0.70)	(0.97)
N	315	315	315	315	315	315	315	315
Cohen's d	0.0400	-0.0700	-0.0100	0.0200	0.0500	0.0400	-0.0400	0

Notes. Estimated differences in outcome measures associated with cohort membership obtained from regression models; p-values in parentheses adjusted for evaluation design clustering.

Table 38: Impacts of AllChild participation at end-point: wide matched sample

	Wellbeing	School Engagement	Parent Engagement	Social Support	AntiSocial
Cohort Membership	-0.0618	0.0561	0.0926	0.0297	0.160
	(0.26)	(0.24)	(0.09)	(0.61)	(0.00)
N	380	410	369	382	383
Cohen's d	-0.120	0.130	0.220	0.0600	0.330
% Change	-6.180	5.610	9.260	2.970	15.99

Notes. Estimated differences in outcome measures associated with cohort membership obtained from regression models; p-values in parentheses adjusted for evaluation design clustering.

Table 39: Impacts of AllChild participation at end-point: narrow matched sample

	Wellbeing	School Engagement	Parent Engagement	Social Support	AntiSocial
Cohort Membership	-0.0584	-0.0126	-0.0381	-0.184	0.0592
	(0.52)	(0.87)	(0.60)	(0.04)	(0.47)
N	141	149	130	138	136
Cohen's d	-0.120	-0.0300	-0.0900	-0.370	0.120
% Change	-5.840	-1.260	-3.810	-18.39	5.920

Notes. Estimated differences in outcome measures associated with cohort membership obtained from regression models; p-values in parentheses adjusted for evaluation design clustering.

4.2 Overall differences using discontinuity design

We now turn to our findings from the discontinuity design. As discussed above, and especially in light of the challenges achieving well-balanced matched comparator groups, this design would generally be expected to estimate an impact that is closer to a truly causal estimate (i.e., the change we would expect for an individual receiving the AllChild programme compared to the genuine counterfactual of them not receiving it, unaffected by selection bias). However, this comes at a cost of being a 'noisier' estimate and, hence, less likely to be statistically significant, not least due to the restrictions on sample size around the attendance discontinuity that we have used.

Furthermore, that sample restriction around the discontinuity means that the impact estimate is not for the full sample of participants in AllChild. Rather, it is an estimate of the impact for young people who are selected onto the programme as a result of just falling below the attendance threshold used as part of the selection process (96% attendance), who would not have been identified otherwise. This is a group who are, by definition, on the margin between whether or not AllChild choose to include them in the programme and, so, may not be representative of impacts among those who are much more likely to be selected.

As discussed above, we only estimate impacts using a discontinuity design for participants in primary schools since initial exploration indicated that the attendance discontinuity did not substantially affect probability of treatment in the secondary school sample.

4.2.1 Mid-point outcomes

Beginning with mid-point outcomes, the findings are reported for the wide discontinuity sample in Table 40 and the narrow discontinuity sample for Table 41, but both tell a similar story. The estimates are in the direction that we would hope they would be and are of a substantial magnitude.

Nevertheless, likely due to the limitations of the discontinuity method with the sample size available, are not statistically significant. This means there is a substantial probability (above the conventional level of 5%) that we would have seen an estimate as large as this by chance under the null hypothesis of no impact.

Table 40: Estimated impacts on mid-point outcomes using discontinuity design — wide sample

	First Stage	Attainment	SDQ
Discontinuity on treatment	0.0841		
	(0.001)**		
Treatment on outcome		1.350	-1.219

		(0.256)	(0.837)
N	805	623	424
F	10.75		
Cohen's d		0.866	-0.180

Notes. Estimated differences in participation associated with being below discontinuity obtained from first stage regression models (First Stage column); estimated differences in outcome measures among those treated due to being below, rather than above, the discontinuity obtained from two-stage least squares regression model; p-values in parentheses adjusted for evaluation design clustering and estimation method. * p < 0.05, ** p < 0.01, *** p < 0.001.

Table 41: Estimated impacts on mid-point outcomes using discontinuity design — narrow sample

	First Stage	Attainment	SDQ
Discontinuity on treatment	0.0667		
	(0.025)*		
Treatment on outcome		0.587	-3.355
		(0.738)	(0.641)
N	590	451	300
F	5.017		
Cohen's d		0.376	-0.508

Notes. Estimated differences in participation associated with being below discontinuity obtained from first stage regression models (First Stage column); estimated differences in outcome measures among those treated due to being below, rather than above, the discontinuity obtained from two-stage least squares regression model; p-values in parentheses adjusted for evaluation design clustering and estimation method. * p < 0.05, ** p < 0.01, *** p < 0.001.

4.2.2 End-point outcomes

Extending then to end-point outcomes (which further reduces sample size, especially as we do not observe end-point outcomes for the most recent cohort), the findings are reported for the wide discontinuity sample in Table 42 and the narrow discontinuity sample in Table 43. Again, it is fair to say that the estimates are in the direction that we would hope they would be but, likely due to the limitations of the discontinuity method with the sample size available, are not statistically significant.

Table 42: Estimated impacts on endpoint outcomes using discontinuity design — wide sample

	First Stage	Attainment	SDQ
Discontinuity on treatment	0.0841		
	(0.001)**		
Treatment on outcome		1.795	-3.757
		(0.323)	(0.651)
N	805	414	439
F	10.75		
Cohen's d		1.127	-0.562

Notes. Estimated differences in participation associated with being below discontinuity obtained from first stage regression models (First Stage column); estimated differences in outcome measures among those treated due to being below, rather than above, the discontinuity obtained from two-stage least squares regression model; p-values in parentheses adjusted for evaluation design clustering and estimation method. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 43: Estimated impacts on endpoint outcomes using discontinuity design — narrow sample

	First Stage	Attainment	SDQ
Discontinuity on treatment	0.0667		
	(0.025)*		
Treatment on outcome		2.009	-7.462
		(0.518)	(0.595)
N	590	274	288
F	5.017		
Cohen's d		1.220	-1.172

Notes. Estimated differences in participation associated with being below discontinuity obtained from first stage regression models (First Stage column); estimated differences in outcome measures among those treated due to being below, rather than above, the discontinuity obtained from two-stage least squares regression model; p-values in

parentheses adjusted for evaluation design clustering and estimation method. * p < 0.05, ** p < 0.01, *** p < 0.001.

Despite the fact that these findings are not statistically significant, it is encouraging to note that, with what we would hope to be the estimates closest to being causal from this, all are consistent with the direction of impact that we would hope for from the intervention.

4.3 Impacts of AllChild programme based on qualitative work

4.3.1 The impact of the programme on short-term outcomes for CYPs

4.3.1.1 Socio-emotional support

Based on interviews with the case study pupils (CSPs), Link Workers, and In-School Leads, we concluded that socio-emotional wellbeing was the primary focus and impact area across the schools involved in the Implementation and Process Evaluation (IPE). The CSPs predominantly viewed AllChild as a source of socio-emotional support. When asked to describe AllChild, CSPs highlighted their experiences with this aspect of support:

> It's like, where you get to talk about your feelings and help you with your emotions (CSP2)
>
> They help us deal with things that happen at home (CSP1)
>
> They talk to you, and they kind of solve some of your problems (CSP5)

Several CSPs specifically mentioned that the programme taught them to recognise various emotional states and ways to address them.

This perception was echoed by the Link Workers and In-School Leads, who also viewed socio-emotional benefits as the programme's primary focus and the most observable impact on the CYPs. When commenting on the programme's impact, Link Workers primarily highlighted improvements in the CYPs' socio-emotional wellbeing. For example, they noted how CYPs felt "safer and more secure within school" (Primary Link Worker, Summer Term 2023 interview). Link Workers across the participating schools consistently emphasised CYPs' increased confidence and improved ability to regulate emotional states as key benefits of the programme.

In-School Leads similarly regarded socio-emotional support as the programme's main intended benefit and a primary motivation for enrolment. They spoke highly of the programme's impact, noting significant improvements in CYPs' socio-emotional wellbeing, confidence, and motivation to engage in learning.

4.3.1.2 Academic support

Throughout the four years of evaluation, academic support did not feature prominently in the accounts of the programme's main benefits. Unprompted mentions of academic support were infrequent among the interviewees, including the CSPs, Link Workers, and In-School Leads. When discussing programme activities, the CSPs focused on games, art-

related activities, and out-of-school trips, with a few mentions of Catch-up Literacy and math sessions In-School Leads acknowledged some provision of academic support ("little pockets of academics" [sic]), but viewed it mainly as secondary, with academic benefits seen as derivatives of improved socio-emotional wellbeing. A Primary In-School Lead summarised this approach well: "The well-being side has to be addressed first and then the academic stuff will happen" (Spring Term 2023 interview). Consistent with this, Link Workers recognised that the support provided was predominantly emotional, with academic assistance offered as needed and limited to English and mathematics. Despite the limited direct academic support, feedback from In-School Leads and Link Workers indicated that improvements in socio-emotional wellbeing positively influenced CYPs' academic performance. Several In-School Leads explicitly acknowledged the direct link between CYPs' socio-emotional wellbeing and their academic performance, as summed up in the following quote: "You can't make progress with their learning until you're in a place of high levels of wellbeing" (Primary In-School Lead, Spring Term 2023 interview).

According to the In-School Leads, improved focus, enhanced emotional resilience, better communication skills, and better relationships with peers enabled CYP to perform better academically. This underscored the importance of addressing socio-emotional needs as a foundation for academic success, even in the absence of direct academic interventions.

4.3.1.3 Impact on children and young people's (CYPs') agency

4.3.1.4 Case study pupils' (CSPs') feedback on agency and choice

Throughout the evaluation, feedback from the CSPs indicated that having a degree of freedom and choice in activities significantly enhanced their enjoyment and engagement levels. The CSPs highlighted the sense of control and freedom as key reasons for particularly liking certain activities.

Art-related activities, such as drawing and Lego building, were among the favourites. These activities were typically deployed during partner sessions or time with Link Workers, offering CSPs a creative outlet. These creative activities allowed CSPs to express themselves "in your way" (CSP2) and "have my creative mind" (CSP1). The nature of these tasks meant that CYPs could make decisions about what to create, how to create it, and when to consider their work complete. This autonomy in decision-making was particularly enjoyed by the CSPs.

The preference for activities that provided creative freedom underscores the importance of agency in fostering positive experiences. By giving CSPs control over their creations, these activities helped boost their confidence and sense of ownership over their work. Moreover, these activities served as a form of self-expression, allowing CSPs to communicate thoughts and feelings that they might have found difficult to articulate verbally. From the perspective of the programme, this was particularly beneficial for CYPs who struggled with traditional forms of communication or those who were dealing with complex emotions.

Some CSPs expressed a desire to have a choice about whether and when to participate in programme activities. They explained that they would not enjoy spending time with Link

Workers or engaging in activities if they felt coerced. This feedback highlighted the critical role of voluntary participation in maintaining CSPs' enthusiasm and positive attitudes towards the programme.

4.3.1.5 Feedback from Link Workers and In-school Leads

Feedback from the Link Workers and In-School Leads indicated an awareness that some CYP struggled with the activities when they were prescribed. They noted that involving CYP more closely in the selection and scheduling of activities could be beneficial. It was argued that some CYP might benefit from a more flexible approach to ensure that their participation in programme activities did not compromise their sense of agency which would also have an impact on the CYP's enjoyment of the programme activities.

Link Workers and In-School Leads acknowledged the difficulty of balancing CYPs' desires for choice with the practical constraints and logistical considerations of the programme. These constraints included timetabling activities and sessions, available resources, and the need to maintain structured and manageable group activities. As a result, it was often challenging to meet each individual child's demands fully.

In summary, this analysis revealed the AllChild programme's primary impact on CYP's socio-emotional wellbeing, consistently reported as the main focus and benefit by CSPs, Link Workers, and In-School Leads. Academic support, while provided, was secondary and viewed as a derivative benefit of improved socio-emotional wellbeing. The findings underscored the importance of voluntary participation in maintaining CSPs' enthusiasm for the programme. The programme positively impacted CSPs' sense of agency, particularly through activities allowing creative expression and choice. However, balancing this desire for agency with practical constraints remained an ongoing challenge. Overall, AllChild's approach of prioritising socio-emotional support laid a foundation for improvements in other areas, including academic performance and personal development. This holistic impact suggests the programme's effectiveness in addressing the diverse needs of the participants.

4.3.2 What is the impact of the programme on long term (end of setting) outcomes for children and young people (CYPs) after they have completed the 2-year programme?

Both CSPs we interviewed after their graduation from the programme shared positive reflections on their experiences with AllChild, fondly recalling the partner activities and the individual time spent with their Link Workers. When asked about the lasting impact of the programme, both indicated that they had learned valuable lessons and skills that they continued to use in their lives post-programme. These lessons predominantly pertained to the socio-emotional domain, such as maintaining patience and showing kindness. Both CSPs expressed that they would participate in the programme again if given the chance.

In general, both pupils seemed to be doing well after completing the programme, reporting improvements in their learning and well-being since graduating from AllChild. The secondary student, in particular, showed a more mature understanding of the

programme's temporary nature and highlighted the importance of handling challenges independently. This suggests that older and younger CSPs might manage the conclusion of the programme differently, although the available data does not allow us to make definitive conclusions in this regard.

It's important to note that the limited sample size of only two post-programme interviews significantly constrains our ability to draw broad conclusions about long-term outcomes. This limitation underscores the need for more comprehensive follow-up studies in the future to better understand the programme's long-term impact across a wider range of participants.

4.3.3 The impact of the programme on the wider community – schools and families

Our aim was to address this question through engagement with families, including parents and caregivers, as well as the In-School Leads, who were best positioned to assess the programme's impact on the wider school community.

Gaining access to families remained a challenge throughout the process evaluation. Consequently, our understanding of the programme's potential impact on the wider community primarily comes from the views of the In-School Leads. They shared reflections on how the programme benefitted the wider school environment.

The In-School Leads provided positive feedback on the programme's impact on schools. It was argued that the following key aspects made the programme particularly valuable:

- The programme's scale and its capacity to reach 30-40 CYP in each school at once.
- The presence of a Link Worker as a dedicated person managing all aspects of the programme, including paperwork and logistics, which minimised the burden on schools.
- The focus of the programme on specific demographics of the intended recipients, as, being at relatively low levels of risk so would not qualify for more urgent forms of intervention, but who still required support. The programme was noted for preventing many CYP from escalating to more severe cases.
- The wide range of partner activities offered to CYP, which would not be accessible to schools on their own.

We consider parental feedback to be critical to understanding the programme's impact on the wider community, specifically families. Reflecting on the challenges in gaining access to parents and caregivers, we understand that key reasons for parents' lack of engagement included time constraints and apprehension about speaking with an outside person as part of a formal process. Similar challenges were faced by the Link Workers, who found it difficult at times to engage with parents or ensure consistent two-way communication. An issue identified early in the evaluation was parents' reluctance to engage with the programme due to social stigma. The Link Workers noted that some families were apprehensive about the programme's nature and goals, with some parents suspecting it was linked to social services, which induced anxiety. As a result, some parents were

sceptical or hard to reach and required extra effort, such as separate meetings with Link Workers, to encourage participation. This suggests the potential value of developing consistent guidelines to assist Link Workers in presenting the programme to families in ways that avoid connotations and associations that may be perceived negatively.

In sum, the evaluation of the AllChild programme's impact on schools and families yielded mixed results. In-School Leads reported positive effects within schools, noting the programme's ability to support 30-40 CYP per school who might otherwise not receive interventions. Strengths included having dedicated Link Workers to manage logistics, focusing on CYP at lower risk levels who still needed support, and offering a range of partner activities. However, the programme's full impact on families remained unclear due to limited parental engagement. Barriers included time constraints, concerns about formal processes, and social stigma. Misunderstandings about the programme's association with social services further hindered parental involvement, underscoring the need for clearer communication strategies. Thus, while the programme appeared to positively influence the wider school community, improving outreach and engagement strategies is crucial for maximising its potential impact on families and the broader community.

5 Understanding the results

Why have we found the results that we have? It is obviously not possible to answer this question comprehensively and with certainty. However, there is more that we can say in order to do so based both on the qualitative work carried out as part of the implementation and process evaluation, and through additional analysis of the quantitative data comparing estimates based on indicators of the extent of participation in the programme.

5.1 Qualitative analysis of AllChild programme delivery

The AllChild programme's theory of change is rooted in providing personalised support to CYP through two key mechanisms: individualised support plans and the presence of Link Workers as trusted adults. The process evaluation focused on these two fundamental aspects due to their crucial role in ensuring the integrity of the programme's design and the success of its implementation.

1. Individualised Support Plans (ISPs): The programme design stipulates that each child receiving the intervention should have an individualised plan of support. These plans are intended to be developed collaboratively by the Link Worker, the In-school Lead, and the CYP themselves. As stated in the AllChild tender proposal, "Each child works with their Link Worker to identify objectives for each term and plan support and activities to reach their goals, meaning their plan is truly personalised." This aspect was chosen for focus because it represents the programme's commitment to tailoring support to each child's specific needs and goals, which is theorised to enhance the effectiveness of the intervention.
2. Link Workers as Trusted Adults: The role of Link Workers as "trusted adults" is integral to the programme's underlying theory. AllChild's formal documentation emphasises this, stating, "Each child has an AllChild Link Worker. They act as a 'trusted adult' and guide them through their programme." This aspect was selected for focus due to its potential significance in building rapport with CYP, which is theorised to be essential for the successful delivery of support and the achievement of positive outcomes.

The study concentrated on these two elements because they appear to be crucial to ensuring the integrity of the programme's design and the success of its implementation. By examining how well ISPs were truly individualised and how effectively Link Workers fulfilled their role as trusted adults (including their presence and availability in the school), the evaluation aimed to assess key components of the programme's theory of change. This focus allowed the study to explore:

1. The extent to which the programme was able to deliver on its promise of personalised support through ISPs.
2. The effectiveness of the Link Worker model in establishing trust and providing consistent support to CYP.

3. How these core elements contributed to the overall impact of the programme on CYPs' outcomes.

By concentrating on these aspects, the IPE sought to provide insights into the mechanisms through which the AllChild programme aims to effect change, and to identify potential areas for improvement in the programme's design and implementation.

5.1.1 Individual Support Plans (ISPs)

The programme design specifies that each child receiving the intervention has an ISP developed by the Link Worker in consultation with the In-school Lead and the CYP themselves. This collaborative approach aims to tailor support to each child's unique needs, preferences, and circumstances.

Our findings demonstrated that CYPs had a limited role in the development of ISPs. Interviews with the Link Workers and In-School Leads revealed that ISPs were partially informed by conversations with the CYPs, focusing on their termly goals, hobbies, and interests. Link Workers combined insights from these discussions with formal assessments of the CYPs needs and risk areas to create ISPs. While CSPs confirmed that discussions took place about their proposed plans of support, particularly regarding partner activities, some expressed uncertainty about why they were chosen for certain activities. This suggested that CYPs' voices were not consistently or sufficiently rigorously integrated into the design of ISPs.

Although CYPs were given choices and their feedback was considered, they were not able to directly select the activities themselves. Several Link Workers and In-School Leads explained that accommodating each child's preferences was impractical due to various logistical factors. These included scheduling constraints, as well as considerations related to gender and year groups, which influenced group dynamics and the most suitable allocation of CYPs to activities. As one Primary Link Worker expressed, "I try to be fair about it, but their voice is not always included."

Remote observation of a partner allocation meeting held in the Spring Term of 2023 confirmed the significance of practical considerations in shaping CYPs' ISPs. Key factors influencing the choice of activities included: * Existing provisions and support in the school * The child's individual interests * CYPs' areas of risk and the goals for them * CYPs' age (with some activities deemed unsuitable for younger CYP) * Previous conduct of the activity in the school (preference for new activities) * Practical and logistical considerations, such as the day and time.

While there was no hierarchical order to these considerations, CYPs' interests did not dominate the discussion; instead, they played a complementary role alongside other factors.

The importance of CYPs' participation in ISP development is underscored by another finding: feedback from CYP demonstrated a direct link between their level of involvement in selecting the activities and their enjoyment. Enhancing CYPs' participation in developing

their plans of support is likely to boost their motivation, enjoyment, and overall benefit from the activities. CYPs' age also emerged as a relevant factor. As noted by one Link Worker, younger CYPs might not have the capacity for in-depth discussions about their goals, suggesting that a more proactive role from Link Workers in designing their plans of support might be beneficial. This reflection underscores the importance of tailoring the approach to designing ISPs to the developmental stage and individual capacities of the CYP, ensuring a balance between practicality and meaningful child participation.

5.1.2 Link Workers as "trusted adults"

Our findings confirm the crucial role of Link Workers as persons of trust, who play a vital role in ensuring that CYPs receive optimal benefits from the programme, particularly in the socio-emotional domain. All Link Workers shared the belief that "building that trusted relationship with the child" was a key aspect of their role. This sentiment underscored their commitment to being more than just facilitators of the programme; they acted as mentors and confidants, crucial for the CYPs' socio-emotional wellbeing.

Our findings indicated that Link Workers had been successful in performing their role as trusted adults. This success was confirmed by feedback from In-school Leads, the CSPs themselves, and the few parents we interviewed. The vast majority of the CSPs reported feeling comfortable discussing personal issues with their Link Workers. They enjoyed spending time with them and expressed a desire for more one-to-one time. Only a few CSPs were reluctant to talk openly to their Link Workers about issues they were struggling with, often due to individual circumstances, which does not detract from the overall positive impact observed.

In-school Leads also confirmed the success of Link Workers in establishing close and trusted relationships with the CYPs. They attributed a significant portion of the programme's benefits to the presence and engagement of Link Workers. They emphasised that CYPs were eager to spend time with their Link Workers and many proactively sought opportunities to do so. It appears that the perception of Link Workers as adults who were separate from the school environment was a significant factor contributing to the trust between CYPs and Link Workers. Both Link Workers and In-school Leads noted that CYPs associated time spent with their Link Workers with "fun" and "doing nice things." This perception was confirmed by the CSPs, who described their Link Workers as creators and facilitators of exciting and enjoyable activities each week.

The continuous presence and availability of Link Workers in the school were highlighted as crucial factors contributing to the programme's success. Both Link Workers and In-School Leads emphasised the importance of this aspect of the programme design. In-School Leads noted that the continuous and long-term presence of Link Workers was a distinctive feature of the programme, which maximised its benefits for CYPs. They believed that having Link Workers consistently available throughout the school year allowed for the progressive building of trusted relationships with CYPs. Link Workers themselves stressed that their ongoing presence and visibility in the school were crucial for fostering trust and rapport with CYPs. This continuous interaction helped Link Workers become a familiar and

reliable part of CYPs' daily lives, which was essential for developing strong, supportive relationships.

The evaluation also revealed some challenges in the Link Worker role. The size of the cohort, which varied between 30 and 40 CYPs, proved challenging for most Link Workers interviewed. They indicated that managing such a large cohort might hinder them from offering the required amount of attention to every child. All Link Workers, except one, acknowledged that smaller groups would allow for deeper relationships with each child. This suggests that while the Link Worker role is largely successful, there may be room for improvement in terms of workload management and resource allocation. In relation to this, some In-School Leads suggested that initiating the programme as early as possible in the academic year and streamlining the induction process could further maximise the programme's benefits. By starting earlier and reducing the time spent on initial setup, Link Workers could engage with CYPs sooner and establish the necessary relationships more effectively.

In sum, we focussed on ISPs and the role of Link Workers as trusted adults as two fundamental mechanisms underpinning the programme's theory of change. The programme design stipulates that each child should have an ISP developed collaboratively by the Link Worker, In-school Lead, and the child. However, the evaluation findings revealed that CYPs had limited involvement in the development of these plans. While Link Workers and In-school Leads considered CYPs' termly goals, hobbies, and interests, the integration of their voices in the design process was not consistently rigorous. Practical considerations, such as scheduling constraints and group dynamics, often took precedence over CYPs' preferences in activity selection. Despite this, the evaluation noted a direct link between CYPs' level of involvement in selecting activities and their enjoyment, suggesting that increased participation could enhance motivation and overall benefit from the programme.

The second key mechanism, Link Workers as trusted adults, proved to be largely successful. Link Workers viewed building trusted relationships with CYPs as a crucial aspect of their role, and this was confirmed by feedback from In-school Leads, the CSPs, and parents. The majority of the CSPs reported feeling comfortable discussing personal issues with their Link Workers and expressed a desire for more one-on-one time. In-school Leads attributed a significant portion of the programme's benefits to the presence and engagement of Link Workers, noting that CYPs eagerly sought opportunities to spend time with them. The continuous presence and availability of Link Workers in schools were highlighted as crucial factors contributing to the programme's success. However, the evaluation also revealed challenges, particularly regarding the size of cohorts managed by Link Workers. Most Link Workers interviewed found it challenging to manage groups of 30 to 40 CYPs, suggesting that smaller cohorts might allow for deeper relationships with each child.

In conclusion, while the evaluation confirmed the importance and general success of these two key mechanisms in the AllChild programme's theory of change, it also identified

areas for potential improvement. These include enhancing CYPs' participation in developing their support plans and considering adjustments to Link Worker cohort sizes to maximise the benefits of their trusted adult role.

5.2 Differences in impact estimates by programme delivery sub-groups

We now turn to further analysis of our quantitative data, exploring a range of outcomes to understand whether there is evidence that the impact estimates vary depending upon the extent of treatment that has been received by participants. We do this by looking at sub-groups of the participants by extent of treatment, especially focussing receipt of AllChild's 'core commitment' (the minimum level of programme time targeted for all programme participants but which cannot always be met for a range of reasons), along with corresponding matched comparator pupils for members of these sub-groups on the basis that these are the most appropriate comparator group for the relevant sub-group.

In order to avoid this section becoming overwhelming (looking across many outcomes and many potential sub-groups) we have focussed on analysis of the narrow matched sample (where there we anticipate lower levels of negative selection bias) and on outcome measures where with larger estimated overall impacts (whether in positive or negative directions).

5.2.1 Primary schools

Building on our overall findings from the mid-point analysis of SDQ outcomes, we explore evidence of difference in these outcomes depending on whether the core commitment was met or not. Building on the overall positive impact estimates, there is a more positive impact for those who meet the core commitment (Table 44 for Peer Problems; Table 45 for Pro-Social). This is consistent with the idea that meeting the core commitment is associated with more positive outcomes in this respect. As with the overall findings, we do not find such evidence in our endpoint analyses of these outcome measures.

Table 44: Impacts of AllChild participation in primary school on Peer Problem SDQ sub-scale scores at mid-point by receipt of core commitment: narrow matched sample

	Not Met	Met
Cohort Membership	-0.0238	-0.471
	(0.93)	(0.00)
N	185	628
Cohen's d	-0.0100	-0.260

Notes. Estimated differences in outcome measures associated with cohort membership obtained from regression models; p-values in parentheses adjusted for evaluation design clustering.

Table 45: Impacts of AllChild participation in primary school on Pro-Social SDQ sub-scale scores at mid-point by receipt of core commitment: narrow matched sample

	Not Met	Met
Cohort Membership	0.154	0.404
	(0.71)	(0.06)
N	185	628
Cohen's d	0.0600	0.160

Notes. Estimated differences in outcome measures associated with cohort membership obtained from regression models; p-values in parentheses adjusted for evaluation design clustering.

Moreover, we continue a note of caution from the overall results that participation in the AllChild programme may not be supporting academic attainment, while noting that this is not emphasised as its primary aim (as discussed by the implementation and process evaluation). In Table 46, we find that primary school participants whose participation in the AllChild programme meets the core commitment see a more negative impact estimate than those whose participation does not meet the commitment. We emphasise, however, that neither of these effects (nor the difference between them) can be considered statistically significant, as well as the caveats around the construction and interpretation of the attainment outcomes across schools discussed in the data section.

Table 46: Impacts of AllChild participation in primary school on attainment scores at end-point by receipt of core commitment: narrow matched sample

	Not Met	Met
Cohort Membership	-0.129	-0.179
	(0.47)	(0.13)
N	244	419
Cohen's d	-0.0800	-0.120

Notes. Estimated differences in outcome measures associated with cohort membership obtained from regression models; p-values in parentheses adjusted for evaluation design clustering.

Conversely, we find more positive impacts of the programme on attendance among primary school participants who receive the core commitment (Table 47). This finding remains statistically insignificant but is directionally consistent with the idea that fully participating in the programme is supportive of this outcome. This distinction in outcomes on attendance is also evident in secondary school participants (Table 48). Again, we see larger impacts on attendance among those who do receive the core commitment (if anything there is an even larger distinction than in primary schools). That said, it is possible

that these apparent differences in effects on attendance are driven by the differences in attendance themselves affecting the ability of AllChild to meet the core commitment. In other words, the ability to meet the core commitment may be being adversely affected by poor attendance.

Table 47: Impacts of AllChild participation in primary school on attendance at end-point by receipt of core commitment: narrow matched sample

	Not Met	Met
Cohort Membership	-0.0729	0.732
	(0.91)	(0.13)
N	261	419
Cohen's d	-0.0100	0.110

Notes. Estimated differences in outcome measures associated with cohort membership obtained from regression models; p-values in parentheses adjusted for evaluation design clustering.

Table 48: Impacts of AllChild participation in secondary school on attendance at end-point by receipt of core commitment: narrow matched sample

	Not Met	Met
Cohort Membership	-2.932	2.698
	(0.10)	(0.07)
N	219	141
Cohen's d	-0.280	0.220

Notes. Estimated differences in outcome measures associated with cohort membership obtained from regression models; p-values in parentheses adjusted for evaluation design clustering.

6 Conclusions

In this report we have reported on UCL's evaluation of the AllChild early intervention programme, reporting findings across our impact evaluation (both based on a matched comparison design, and a discontinuity design) and our implementation and process evaluation. These two aspects focussed on answering slightly distinct, but overlapping, sets of questions about AllChild's work. To conclude, we draw out a number of lessons about AllChild's work, its impact, and the challenges faced in doing this evaluation work.

6.1 No broad impact on available quantitative measures

We were unable to identify consistent evidence of impact associated with participating in the AllChild programme compared to the estimated counterfactual on the available quantitative measures. However, we should temper this conclusion in a couple of ways.

First, the challenges that we noted in setting out the evaluation design. While we have successfully been able to identify comparison groups that are a substantial improvement on attempting to evaluate the impact of AllChild simply by comparing participants with non-participants, we caution that observable (and, likely unobservable) differences do remain between the treatment and comparison groups. This is the case even where we have taken quite an aggressive approach to identifying well-matched comparison groups at the expense of sample size – ultimately young people selected into AllChild are extremely different to their peer group and, as a result, it is extremely difficult to identify a suitable number of truly comparable individuals as a comparison group. Remaining differences between participants and matched comparison groups likely imply continued negative selection bias in our estimates. In that light, the lack of positive estimates may be caused by this issue, especially as the estimates from our discontinuity design — which we would anticipate being closer to causal estimates — tell a more positive picture. However, the estimates from the discontinuity design, while quite substantial, are rather imprecise: even with larger estimates they are not statistically distinguishable from no impact, and are limited to primary schools.

Second, there are positive spots evident. These include encouraging findings on domains of the SDQ and findings that suggest there may have been more positive impacts where the programme is delivered in full, for example, more positive impacts on attendance evident in both primary and secondary contexts when focussing only on participants who received AllChild's core commitment.

As such, we caution that this report should not be interpreted as strong evidence of no impact, but rather lack of strong evidence in favour of impact on these particular measures. Challenges inherent in the evaluation design that was jointly deemed feasible (chosen to include a number of features valuable to addressing a number of additional research questions but, ultimately, negatively affecting the overall design) have contributed to this outcome, along with attrition rates (i.e., missing outcome measures, especially for endpoint outcomes) to a much greater extent than is typically needed for

robust evaluation evidence. We would encourage AllChild and other organisations seeking to continue their evaluation journey to:

- plan to ensure outcomes data can be collected from as many participants and members of the comparison group as possible;
- prioritise an evaluation design that delivers a well-balanced comparison group even if this means less flexibility in other aspects of the design.

6.2 Focus on socio-emotional support

To the extent we saw more promising findings in the impact evaluation, these were more evident on socio-emotional-related outcomes, such as some domains of SDQ at midpoint. This triangulates with findings from the implementation and process evaluation, which found that the programme demonstrates a strong emphasis on socio-emotional support, aligning well with schools' primary motivation for enrolment. Findings indicate success in improving recipients' socio-emotional well-being. The CSPs greatly enjoyed the programme, particularly valuing:

- Individual time with Link Workers
- Partner activities
- Out-of-school trips
- Open-ended and creative activities which offer opportunities for individual choice

To further enhance CYPs' enjoyment and engagement, we recommend increasing their agency within the programme. This can be achieved by:

- Providing options: Establish a clear process for CYP to express preferences and make choices about partner activities, balanced with logistical considerations
- Regular check-ins: Schedule periodic check-ins with CYP to gather input about their experiences and adjust plans where possible
- Transparent communication: Explain decision-making processes to CYP, helping them understand any constraints while actively involving them in finding solutions.

6.3 Strengthening the role of Link Workers

Link Workers play a pivotal role as trusted adults in supporting CYPs' socio-emotional well-being. The evaluation found that Link Workers have been highly successful in establishing close relationships with the CYPs. To further enhance this aspect, we recommend:

- Early programme initiation: Start the programme as early as possible in the academic year to allow Link Workers to begin relationship-building sooner.
- Streamlined induction process: Shorten or optimise the induction process to increase direct engagement time between Link Workers and CYP.

- Optimal Link Worker-to-child ratio: Consider reducing cohort sizes or providing additional support to enable more personalised attention for each child.

6.4 Ensuring continuous and consistent engagement with families

Parental involvement remains a challenging aspect of the programme. To maximise the wider benefits on families and communities, we recommend:

- Regular communication: Maintain diverse, consistent channels of communication with parents to keep them informed about their child's progress and programme activities.
- Addressing barriers: Identify and mitigate common obstacles to parental engagement, such as time constraints and apprehensions about formal processes.
- Inclusive approach: Develop strategies to enhance programme accessibility and appeal to parents, including addressing social stigma and misconceptions about the programme's nature and goals.

6.5 Conclusion

We appreciate that this evaluation has not delivered the conclusive positive impact estimates that any organisation would hope to achieve in an evaluation of their work, not least due to the challenges experienced in identifying a comparable comparison group within constraints due to the nature of the programme.

Nevertheless, we note that it there are a number of important lessons to be taken from this research, both in terms of informing the future evaluation journey of AllChild and in understanding potential strengths and weaknesses of the programme in its current form. These include a number of specific recommendations that we suggest would help the programme to build upon its existing strengths and address key areas for improvement, ultimately enhancing its impact on children and young people, families, and communities.

7 Appendix: Codes and themes arising from qualitative data collected

Codes	Main themes
socio-emotional academic socio-emotional vs academic new experiences CYP's feelings confidence enjoyment/fun helping soft skills subject areas measuring impact tangible vs intangible impact CYP's agency choice	RQ1: The impact of the programme on short-term outcomes for CYP
programme's scale and reach range of support preventing escalations expanding horizons relieving pressure on families engaging with parents presentation of programme attitudes towards programmes stigma understanding the programme approach to communication w families consent gathering/paperwork	RQ3: The impact of the programme on the wider community, i.e. schools and families
building rapport with CYP plans of support vs CYP's needs CYP's choice and voice Link Workers' availability and presence Link Workers' role for CYP Link Workers' status in school mediating between schools and families CYP's feelings about Link Workers trust boundaries Link Workers' capacity ending activities	RQ4: Theory of Change, including: Individualised plans of support Trusted adults

ucl.ac.uk/ioe/cepeo