



Evaluation of the Reboot programme: Feasibility study report

November 2022

Contents

Contents	1
Executive summary	2
1. Introduction	4
1.1 Programme overview	4
1.2 Programme description	4
1.3 Research objectives	5
2. Methods	7
2.1 Explore potential evaluation designs and associated requirements	7
2.2 Identify and refine of possible outcome measures	7
2.3 Assess the requirements and feasibility of different evaluation design options	8
2.4 Develop evaluation design recommendations	8
3. Considerations for outcome measures	9
3.1 Principles and criteria used to guide our selection outcome indicators	9
3.2 Data sources	11
3.3 Recommendations for selecting outcome indicators	15
3.4 Proposed outcome indicators	17
4. Considerations for evaluation design	21
4.1 Key considerations	21
4.2 Sample size	25
4.3 Comparison group	28
4.4 Data availability	28
4.5 Feasibility	31
4.6 Lessons from the feasibility assessment	38
5. Recommendations and next steps	41
5.1 Recommended outcome indicators	41
5.2 Recommended evaluation design	42
5.3 Next steps	44
Appendix A. Additional figures and tables	46
Appendix B. Testing relevance and sensitivity: outcome indicators	50
Appendix C. Power calculations	54

Executive summary

Context

The Reboot programme aims to support care-experienced young people into work, education or training. The programme targets young people aged 16-25 who are care-experienced and looked after by statutory care in the West of England Combined Authority (WECA) or North Somerset Council, are not currently in education, employment or training (NEET), or are at risk of not being in employment, education or training.

This report sets out the findings of a feasibility study conducted by the Behavioural Insights Team to determine the most appropriate design for the impact evaluation of the third iteration of the programme, Reboot III, intended to begin in 2023.

Methods

The feasibility study was conducted between November 2021 and October 2022, and work took place in the following stages:

1. Explore potential evaluation designs and associated requirements
2. Identify and refine possible outcome measures
3. Assess the requirements and feasibility of different evaluation design options
4. Develop evaluation design recommendations

The feasibility study was conducted in parallel with a process study and participant tracing aiming to generate evidence on the delivery of the Reboot II programme and how it supports young people to achieve EET outcomes.

Recommendations for outcome indicators

Based on our review of available data and our assessment of possible indicators against our criteria of *relevance* to the programme's intended outcomes, the *feasibility* of obtaining data for both programme participants and a potential comparison group, and *sensitivity* (i.e., ability to capture variation across young people and time), we **recommended the following principles in selecting outcome indicators** for the impact evaluation:

1. Select indicators that **focus on EET outcomes**
2. Select indicators that **can be constructed based on data held by Local authorities (LAs), and enriched with data from other sources**
3. Select indicators that **focus on the latter stages of the programme participant journey**

In line with our indicator selection criteria and the principles above, we recommended the following outcome indicators for the impact evaluation¹:

¹ The final impact evaluation design includes secondary outcome measures derived from HMRC data that was agreed following the delivery of the feasibility study. These are; employment status; time in employment; and earnings.

- **EET status (primary outcome):** A binary outcome variable that indicates whether the young person is engaged in part or full time employment, education or training for the majority of the latter stages of a two year participant journey (last 6 months).
- **Position on EET scale (secondary outcome):** A number indicating the level achieved on a predefined EET scale by the young person as a rough guide to the progression away from risk of NEET. The scale will have a fixed number of levels, where each EET outcome is assigned a certain level, with more desirable outcomes being scored higher. We also recommend exploring the option of having multiple secondary outcomes, each focused on one of the dimensions of EET.

While these indicators can be based strictly on LA data if it is sufficiently complete and reliable for both a treatment and control group, the feasibility of a rigorous evaluation would likely be significantly higher if data from HM Revenue and Customs (HMRC) and the Department for Education (DfE) are accessed and used to enrich LA data. However the availability of DfE data specifically for measuring outcomes is likely to fall outside the current planned evaluation window.²

Recommendations for evaluation design

To guide our thinking in identifying and assessing the feasibility of different evaluation design options, we **considered four factors**:

1. Sample size
2. Potential comparison groups
3. Data availability
4. Implementation feasibility

Based on our review and assessment of the factors above, we recommended that **the evaluation design that best balances rigour and feasibility is a randomised controlled design, in which participants for both the Reboot group and comparison group are sourced within WECA and North Somerset Council**. Given the anticipated challenges securing detailed data from non-participants in LAs outside WECA and North Somerset, it was likely to be more challenging to construct a sufficiently large and valid comparison group for a quasi-experimental design. From discussions with LAs and based on modelling estimates discussed with 1625ip, our assessment is that **randomisation was possible to implement and would allow for a sufficient sample size**.

Overall, we have been conservative in our assumptions, but it is worth noting that our best assumptions, based on the information available, have been used to determine the potential sample size for the evaluation and the estimated effect of the programme. Given this, there is a need to ensure recruitment targets are reached to mitigate the risk that the trial will not be sufficiently powered to detect an effect, and this has been considered by YFF alongside the mitigation strategies discussed in this report.

² At time of writing, the approach to the potential inclusion of LEO data had not been agreed. Since the delivery of the feasibility report, the potential for using LEO has increased. This has been reflected in the trial protocol. Should LEO data become available, these data will be used to reconstruct the outcome measure and replace LA outcome data for the analysis of primary outcomes.

1. Introduction

1.1 Programme overview

The Reboot programme works with young people aged 16-25 who are (or have been) looked after by statutory care services in the west of England, and provides them with coaching support for up to three years to help them obtain and sustain employment, education and/or training (EET). The frequency and type of support varies, but it is based on a youth version of Acceptance and Commitment Therapy, called DNA-V. DNA-V is a psychological intervention that uses acceptance and mindfulness, together with commitment and behaviour change strategies, to improve psychological flexibility. In turn, this helps the young person to stay in contact with the present moment regardless of unpleasant thoughts, feelings or bodily sensations, while choosing behaviour and action based on the situation and personal values.

The programme is delivered by 1625 Independent People (1625ip), a homelessness charity based in the south west of England, with funding from the Youth Futures Foundation (YFF) and the West of England Combined Authority (WECA). YFF is an independent What Works Centre for youth employment established to improve employment outcomes for young people from marginalised backgrounds. WECA is a combined authority within the west of England, consisting of the local authorities (LAs) of Bristol, South Gloucestershire and Bath & North East Somerset.

The first iteration of the programme, Reboot I, has been running since August 2018 and is funded through the Department for Education's (DfE's) Social Care Innovation Fund through a social impact bond. Reboot I worked with 237 care-experienced young people, all not in education, employment or training (NEET) or at risk of NEET. The second iteration of the project, Reboot II was operational between 2021 and 2022.

The following sections of this report set out the findings of a feasibility study conducted by the Behavioural Insights Team (BIT) to determine the most appropriate design for the impact evaluation of the third iteration of the project, Reboot III.

1.2 Programme description

BIT organised a workshop with staff from 1625ip and the involved LAs to get an understanding of the Reboot programme activities, what they aim to achieve, and how. Based on these learnings and on subsequent feedback, we created a programme theory of change. This covers the activities undertaken as part of the Reboot programme, the outputs expected from these activities, the short- and long-term outcomes these outputs are meant to achieve, as well as the target audience, assumptions, and the longer-term benefits and cost-savings.

The theory of change in Figure 1 sets out that upon entry to the programme, there are a range of support activities for young people, including developing an action plan, regular

one-to-one sessions with a coach, EET support activities, coaches responding to ad-hoc calls and requests, and support with finances. There are also a range of support and professional development activities for coaches.

These activities are intended to lead to a range of outputs, such as the young person having a better understanding of their abilities and goals, being better able to apply for employment and education opportunities, having a better understanding of healthy relationships and how to build them, and being able to access resources and financial support.

The outputs identified are expected to lead to a range of outcomes. In the short-term, these intended outcomes are: obtaining a place on an educational course (if relevant); increasing the young person's psychological flexibility; improving their relationships and support network; and ensuring their basic needs are satisfied. In the long-term, these should then lead to obtaining educational qualifications and employment, a greater physical and mental wellbeing, and obtaining, sustaining and progressing in meaningful employment.

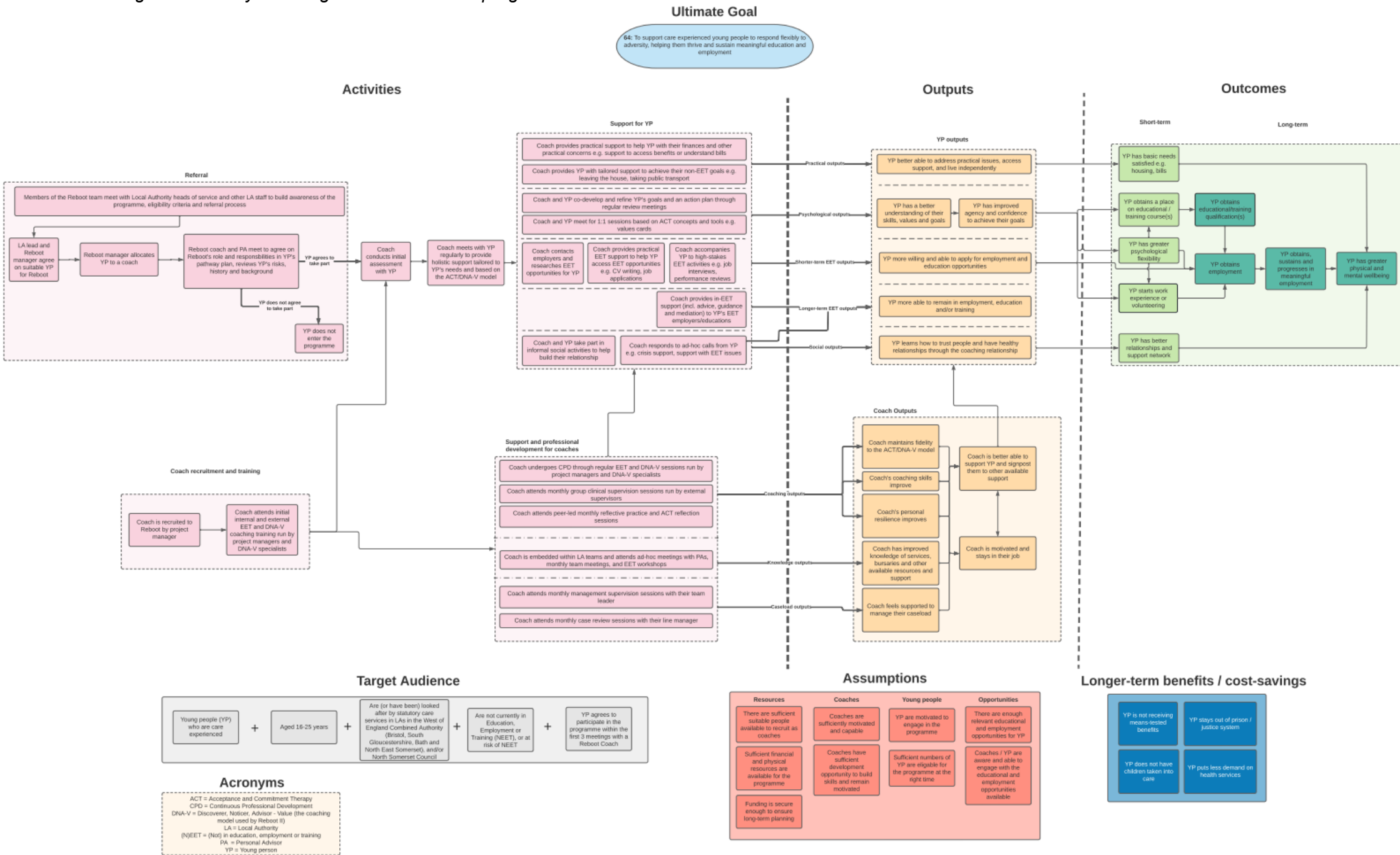
1.3 Research objectives

The objective of this feasibility study was to identify an evaluation design that best balances rigour and feasibility in answering the research questions jointly identified with 1625ip, participating LAs and YFF.

After a description of the methods followed for this feasibility study, the third section in this report focuses on determining the most appropriate outcome measures for impact evaluation, including whether the programme helps young people obtain educational qualifications and employment, and if it helps them to sustain and progress in employment. The fourth section examines the different evaluation design options scoped throughout the study, and presents the programme and data requirements of each option, and the associated trade-offs between different options. Finally, recommendations for the most appropriate evaluation design are described in the fifth section.

BIT is conducting a process and participant tracing study in parallel with this feasibility study to examine the earlier stages of the theory of change, and identify potential refinements to programme design or implementation before the launch of an impact evaluation.

Figure 1. Theory of change for the Reboot II programme



2. Methods

In this section of the report, we set out the methods used to conduct this feasibility study. The study was conducted between November 2021 and October 2022, and was structured in the following stages:

1. Explore potential evaluation designs and associated requirements
2. Identify and refine possible EET outcome measures
3. Assess the requirements and feasibility of different evaluation design options
4. Develop evaluation design recommendations

This feasibility study built upon a scoping phase from July to October 2021, during which we developed a detailed programme theory of change jointly with the 1625ip team and representatives from participating LAs. This theory of change enabled us to build a shared understanding of the programme and its activities, to clearly define its objectives, and to create a visual representation of the causal chain between its activities and outcomes. This diagram acts as a framework for the evaluation design and is used as a key input for many of the stages of work outlined below.

The feasibility study was conducted in parallel with a process study and participant tracing, which aimed to generate evidence on the delivery of the Reboot II programme and how it supports young people to achieve EET outcomes.

2.1 Explore potential evaluation designs and associated requirements

The objective for this stage was to build a shared understanding of the different evaluation design options, the associated programme and data requirements, and the trade-offs between different options.

Drawing on our understanding of the programme and evidence needs, we created an initial list of evaluation design options, setting out the requirements for each and the associated risks, which we discussed with 1625ip and YFF. While we made initial recommendations about suitable designs based on an assessment of their rigour and feasibility, we provided 1625ip with a more detailed description of their requirements to facilitate further in-depth discussions on the feasibility of each design.

2.2 Identify and refine of possible outcome measures

The objective of this stage was to identify possible outcome indicators for the impact evaluation, and to prioritise them based on their relevance to the programme's intended outcomes, the feasibility of obtaining data for both programme participants and a potential comparison group, and their sensitivity (i.e., ability to capture variation across young people and time).

As data availability is a key constraint, we started by identifying different possible sources of data for both Reboot participants and non-participants, including data held by LAs and national administrative datasets. Following discussions with stakeholders and a review of these data sources, we set out potential outcome measures that could be feasible to construct based on the data that could be accessed, and assessed their relevance and sensitivity through a series of discussions with 1625ip, review of past programme data, and initial power calculations. We developed our recommendations through an iterative process and discussions with 1625ip and YFF.

2.3 Assess the requirements and feasibility of different evaluation design options

The objective of this stage was to determine the feasibility of estimating the impact of the programme on the proposed outcome indicators for each of the evaluation design options considered. The key considerations that we analysed to assess the feasibility of each design were: sample size; data availability; existence of potential comparison groups; and implementation feasibility.

A key activity during this stage was to conduct power calculations for the proposed outcome indicators. Using data from Reboot I to inform our assumptions, we identified a set of scenarios for the possible values of the parameters influencing power (e.g., sample size, average outcome value in the comparison group) and conducted power calculations for our selected outcome measures to identify the range of effect sizes that the evaluation would be able to detect for these scenarios. We presented a series of power calculations to 1625ip and YFF to get feedback on the different scenarios used and how the effect sizes estimated compared to expectations for the programme. Additionally, to understand the potential for increasing the Reboot cohort size in line with the sample size scenarios used in the power calculations, we conducted a 'recruitment deep-dive' where we mapped the recruitment process in detail and identified key dependencies that directly affect cohort (and thus possible sample) size. Finally, we obtained aggregated data from LAs currently implementing the programme and consulted publicly available data to refine our estimates of the number of care leavers that could be eligible for the next cohort of the programme in these LAs.

Building on the review of possible data sources and through discussions with Bristol City Council, we also explored the feasibility of obtaining the data required and of constructing a valid comparison group for each design option.

2.4 Develop evaluation design recommendations

The objective of this stage was to articulate recommendations on the best-suited evaluation approach based on the insights from the previous three stages. These recommendations are outlined in this report.

The remainder of this report sets out our findings from each stage of the work, with an explanation of our rationale for decisions made, our design recommendations, and proposed next steps.

3. Considerations for outcome measures

An essential early step in our feasibility study was to define the outcome indicator(s) that the evaluation would use to estimate impact. These indicators should capture the outcomes in the programme theory of change.

3.1 Principles and criteria used to guide our selection outcome indicators

Summary: in selecting outcome indicators, we followed two principles:

1. **Select** indicators that strike the best **balance between relevance, feasibility and sensitivity**; and
2. **Limit the number of outcomes** to maximise our ability to make impact claims.

In defining outcome indicators for an impact evaluation, we followed the two key principles described below.

Principle 1: limit the number of outcomes to maximise our ability to make impact claims

While outcome indicators have previously been defined for the Reboot programme, these were defined in the context of the social impact bond (SIB) conducted around Reboot I, and were conceptualised as targets that, when reached by programme participants, would trigger payments. These target indicators are described in Table A1 at Appendix A. Given the purpose of these indicators as triggers to results-based payments, and the multitude of possible outcomes to be achieved throughout the duration of the Reboot programme, having a high number of outcome indicators was appropriate for previous SIB-focused evaluation efforts of the programme. In contrast, for an impact evaluation, it is important to minimise the number of outcome indicators that will be examined in order to maximise our ability to make impact claims. When estimating the impact of the programme on multiple outcomes, the risk that (at least) one of the results is a false positive increases. To correct for this, the threshold for what is considered a statistically significant finding has to be increased. A common method is the Benjamini-Hochberg step-up procedure. Using this method, the more outcome variables there are, the lower the p-value has to be in order for a result to be deemed statistically significant.

To have the highest chance of finding a robust significant result, we thus aimed to focus only on the most important outcomes in the theory of change, and to define fewer outcome indicators.

Principle 2: select indicators that best strike a balance between relevance, feasibility, sensitivity

We assessed and prioritised possible outcome indicators based on three criteria:

1. **Feasibility:** we can obtain reliable data on the indicator for participants and a possible comparison group.
2. **Relevance:** the indicator captures the real-world behaviour we care about as closely as possible.
3. **Sensitivity:** the indicator captures a high amount of information (variation across time and individuals).

To identify an initial set of possible indicators, we focused on feasibility first, by aiming to understand what potentially relevant data is already being collected on young people eligible for the Reboot programme. We also researched whether and how we may be able to access this data.

There are trade-offs between these criteria across possible indicators related to Reboot target outcomes. We identified and discussed with 1625ip the following challenges related to these trade-offs:

- **Capturing the ‘meaningfulness’ of an EET activity.** A young person moving from one job to another that is more in line with their aspirations and goals would be a positive outcome. An outcome indicator that can capture such a change would score high on relevance. An indicator that could measure the extent of this alignment would also score high on sensitivity. However, as meaningfulness is often subjective and difficult to capture, such an outcome measure would likely not score high on feasibility.
- **Capturing whether the young person can sustain an EET activity.** A young person’s EET status can change frequently. This can be for negative reasons (e.g., dropping out of school or losing a job), or positive reasons (e.g., finishing a degree or finding a better job). In general, if the young person is out of EET for a positive reason, this usually lasts for a shorter duration than if it is for a negative reason. However, if we measure EET status at only one moment in time, there is a risk that this occurs while the young person is in between two activities. An outcome measure measuring EET status over a longer time period would reduce the risk of wrongly ‘penalising’ a young person for briefly being in NEET for a ‘positive’ reason, or ‘rewarding’ a young person for an EET activity they do not sustain. Such an outcome measure would thus be more relevant, but would also likely be less feasible, as it would require data on EET status at multiple time points within a short period of time.
- **Accounting for progression over time.** Young people all have different starting points, and success in terms of EET can mean something different for each. Designing an outcome indicator that accounts for this fact and capture progression granularly would result in high relevance and sensitivity, but would also result in a low feasibility of obtaining the required data.
- **Assessing the relative value of different categories of EET activities.** To minimise the number of outcome variables (per principle 1), we aspire to have a primary outcome variable that captures most positive outcomes, in particular in employment, education and training. Doing this forces us to determine the relative value of these different EET activity categories. Equating the value of different categories of EET activities would make an outcome indicator more feasible, however it would also make it less sensitive.

In considering possible outcome indicators, we aimed to navigate these trade-offs to identify the indicator(s) that best balanced feasibility, relevance and sensitivity.

3.2 Data sources

Summary: we identified various datasets that could provide information relevant to the outcomes in the programme theory of change.

- Of these, **the most feasible and relevant dataset is collected by LAs on care leavers** in their area, and we have reviewed anonymised LA data to assess its quality.
- A limitation of LA data is that the statutory requirement to collect it ends when care leavers turn 21. YFF have negotiated an agreement for LAs to collect data for relevant young people beyond that age for the purposes of the evaluation.
- **HM Revenue and Customs and the Department for Education also collect relevant data** that may be accessible and could be used to increase the sensitivity and relevance of the outcome indicators³.

To understand the range of possible indicators feasible, we identified various datasets that could provide information relevant to the outcomes in the programme theory of change. Most of the information available in these datasets relate to EET outcomes. They cover only a small amount of information related to some non-EET outcomes such as access to basic needs, and no information related to other non-EET outcomes, such as care leaver wellbeing, psychology flexibility and relationships. The datasets we identified come from three different sources described below.

Based on this review of existing datasets:

- LA data on care leavers was identified as highly relevant and feasible to access given the involvement of the LAs in implementing the programme. 1625ip data is similarly relevant and feasible to obtain, however it is only available for Reboot participants, and as such could not be used as a data source for a possible comparison group.
- HM Revenue and Customs (HMRC) and the Department for Education (DfE) data was identified as containing relevant information related to employment and education outcomes, respectively, and as potentially accessible through formal requests to the appropriate government departments. However DfE data for measuring outcomes will not be available within the timeframes of the evaluation; it may however be used to collect personal characteristics of participants to increase the power of the trial.

1625ip

1625ip collects data on a range of indicators for its participants. The purpose of this data collection is primarily for the SIB targets. They thus record education outcomes and

³ Since the delivery of this report, HMRC data access has been confirmed and this has enabled us to include new secondary outcome measures in the design of the impact evaluation. These are: employment status; time in employment; earnings. As previously mentioned, access to LEO data and an approach to the analysis of these data has also since been agreed.

progression, employment and earnings data (through data obtained from HMRC, see below), and volunteering and work experience data. They also collect information on non-EET outcomes such as their accommodation and whether they feel safe, as these outcomes are linked to payments.

The type and amount of data 1625ip collects is flexible, and we are for example assisting them in updating the indicators related to non-EET outcomes that are being collected. However, as they can only collect data on Reboot participants, it is not of direct use for the impact evaluation, though it could be used to validate data from other sources.

LAs

Every LA is required to collect data on care leavers between the ages of 17 and 21 on a number of indicators for what is called the 'annual children looked after return'. The statutory requirement is for LAs to return this data annually, however according to information provided during conversations with Bristol City council, LAs may collect data as frequently as every 2 months, although in practice it is likely to be less.⁴

The data collected covers a range of demographic characteristics, living situation, and crucially, EET activities the young person is engaged in. This information is entered manually into a database using a data management solution specifically developed for this purpose. Information is usually gathered through direct contact with the care leaver. The availability and quality of the data is thus dependent on the extent to which LAs consistently stays in contact with the care leavers, and on how accurately the information is collected and entered at points of contact with care leavers.

An important point to note is that LAs only have a statutory requirement to collect data on care leavers until the age of 21, whereas 1625ip generally work with Reboot programme participants between the ages of 16 and 25⁵. Thus, not all Reboot participants will be covered by this statutory requirement. An exception applies for young people with an education, health and care (EHC) plan⁶, for whom LAs must keep collecting relevant data until they are 25, and those who are in full time education. Among the Reboot I cohort, 54% of participants were 21 or older two years after enrolment in the programme; a similarly large proportion of participants would likely be 21 or above for later cohorts.

We have had conversations with all four LAs about whether and how data could be collected for other care leavers above the age of 21. YFF has reached an agreement for the LAs to collect data approximately every three months for every young person - including those above the age of 21 - during their first 18 months in the treatment or control group, and then approximately every two months between months 18 and 24.

⁴ Bristol provided information showing that they manage this for around 80% of care leavers. Another local authority provided data that suggests an average number of touch points per care leaver of around 3 per year.

⁵ If a young person starts Reboot at an older age, then there is a possibility that they can be older than 25. However, the oldest starting age in Reboot was 23 (for just over 1% of all referrals). Thus, it is unlikely any of the young people will be over 25 during the trial

⁶An EHC plan is for young people aged up to 25 who need more support than is available through special educational needs support. We don't know the percentage of care leavers who have an EHC plan, but it is a minority. This information must be reported to the Department for Education's National Client Caseload Information System (NCCIS). The required indicators are not exactly the same as what is required for all care leavers up to 21, but it contains the ones we are interested in.

HMRC

HMRC collects earnings data on all employees directly from employers, primarily for income tax and National insurance purposes. This data captures whether a person is employed, how many hours (categorised) they work, and their earnings. It is updated in real time (employers must generally send pay details to HMRC on or before the time they pay their employees) and can be shared monthly. As the data is used for important purposes such as calculating income tax and national insurance contributions, the quality and completeness of the data is expected to be very high.

HMRC data is currently used to check and complete some of the SIB indicators for Reboot I. The organisation in charge of the evaluation of Reboot I as part of this SIB, Bridges Outcomes Partnerships, has a data sharing agreement in place with HMRC to obtain earnings data for the young people in the Reboot I cohort. **1625ip and BIT have set up a similar data sharing agreement for Reboot II and Reboot III.**

Department for Education (DfE)

The DfE has rich datasets containing education-related indicators, including highest level of diploma achieved, enrolment status and absences. The main databases relevant to Reboot outcomes are the National Pupil Database (NPD, which covers schools), the Individualised Learner Record (ILR, which covers further education) and Higher Education Statistics Agency's database (HESA, which covers higher education). This data is updated 1-3 times a year, sometimes with delays of over a year. The process of getting a data sharing agreement with the DfE would be similar as with HMRC (and does not need to be done separately for each dataset); however unlike with the HMRC, there is no precedent yet to get this data for Reboot participants. We envisage accessing this data to enhance the data that we have on participants' background (to increase the power of the analysis). However the availability of this data for measuring educational *outcomes* is likely to fall outside the current planned evaluation window.

We also considered DfE's longitudinal educational outcomes (LEO) dataset, but determined that it would be unlikely to be applicable at the analysis stage of this evaluation due to the lag in outcome data, which would result in substantial delays in reporting. **However, by collecting UPNs and Anonymous Pupil Reference numbers as we have agreed to do, we expect LEO outcomes to be viable in the longer term should YFF be interested in exploring these⁷.**

Table 1 describes the type of information relevant to programme outcomes that are captured in each of the three last data sources described above (we do not include 1625ip data as it would not be available for young people not participating in the Reboot programme). Note that this list is indicative rather than exhaustive.

⁷ As previously set out, the use of LEO data has since been agreed for analysis of the impact evaluation primary outcomes.

Table 1. Available indicator captured in potential data sources

Outcome category	Information covered
Employment	<ul style="list-style-type: none"> ● Normal hours worked (less than 16, 16 – 24, 24 – 30, 30+) ● Taxable pay for a given period ● Employment start and end date ● Young person is engaged in full time (>17h) or part time (<17h) employment or training ● Young person is engaged in a full time (>17h) or part time (<17h) apprenticeship ● Reason for not being in EET (illness or disability; pregnancy or parenting; other circumstances)
Education	<ul style="list-style-type: none"> ● Young person is engaged FT or PT in [higher education] / [education other than higher education] ● Programme start and end date ● Programme type / level of study ● Highest qualification on entry in higher education ● Learning outcome ● Student load (Full time equivalence, only HE) ● Exclusions (type and length) at maintained (state-funded) schools ● Authorised absences (number of sessions) at maintained (state-funded) schools ● Unauthorised absences (number of sessions) at maintained (state-funded) schools
Training	<ul style="list-style-type: none"> ● Young person is engaged in full time (>17h) or part time (<17h) employment or training ● Young person is engaged in a full time (>17h) or part time (<17h) apprenticeship
Other	<ul style="list-style-type: none"> ● Accommodation type ● Strengths and difficulties questionnaire score (only for <18s)

Data source and frequency: ● HMRC (every month) ● LAs (every 2 months) ● DfE (every 4 months or year)

3.3 Recommendations for selecting outcome indicators

Summary: based on our review of existing datasets and our assessment of possible indicators against our selection criteria, we recommended selecting outcome indicators that:

1. **Focus on EET outcomes;**
2. **Can be constructed based on data held by LAs**, and enriched with data from HMRC and/or DfE;
3. **Focus on the latter stages of the programme participant journey.**

Given the indicator selection criteria and data sources described above, we made three recommendations below regarding the approach to define outcome indicators for the impact evaluation of the Reboot programme. We discussed these recommendations with 1625ip, and used them as the basis for our proposed outcome indicators.

Recommendation 1: select indicators that focus on EET outcomes

The theory of change's long-term outcomes focus on the behaviours of obtaining, sustaining and progressing in meaningful employment ('EET outcomes'), and on greater physical and mental wellbeing ('non-EET outcomes'). Indicators capturing both EET and non-EET outcomes are thus highly relevant to an impact evaluation of the Reboot programme. However, the feasibility of capturing EET outcomes for both Reboot participants and a possible comparison group is much higher than collecting non-EET outcomes. Further, the primary long-term goal of the programme is to help young people thrive and sustain meaningful employment. Although improving their physical and mental wellbeing is an important outcome on its own, it is also expected to be an intermediary outcome contributing to improved EET outcomes, and would thus be partially captured through EET outcome indicators as well.

Therefore, we recommended focusing on EET outcomes for the impact evaluation indicators. That being said, non-EET outcomes are also relevant to the Reboot programme, and in addition to primary analysis focused on EET outcomes, we recommended analysing Reboot participant data on non-EET outcomes to understand the mechanisms of impact and the relationship between EET and non-EET outcomes.

Within EET outcomes, we recommended defining an outcome indicator that captures both employment and education outcomes, rather than a separate one for each. Having only one primary outcome indicator increases our ability to make impact claims, both because it minimises the threshold required to claim a statistically significant impact and because it ensures that the indicator is relevant to the full sample, rather than limiting analysis to a subset of the sample (e.g., excluding young people in education from the analysis for employment outcomes), which may also invalidate the analysis.

Recommendation 2: select indicators that can be constructed based on data held by LAs

The LAs are the most feasible source of data. If the evaluation can be conducted based solely on their data, we do not have to rely on data sources that are still more uncertain. Selecting indicators that can be constructed based on this data would thus increase the

feasibility of a rigorous impact evaluation.

Additionally, 1625ip collects additional useful data as well, but only on Reboot participants. While we thus cannot use this data to construct an outcome indicator, it can be used to validate the data available on Reboot participants from other sources⁸.

Recommendation 3: select indicators that focus on the latter stages of the programme participant journey

A young person's EET status is variable and likely to change multiple times over the evaluation period, which is currently assumed to be two years. The Reboot programme aims to support young people such that they are able to sustain EET activities after participation in the programme. Therefore, EET activities in the early stages of a young person's journey in the programme should contribute to improving their EET outcomes by the end of the programme and beyond. EET activities in the early stages of the participant journey are not on their own a positive programme outcome if the young person is no longer in EET by the end of the programme. For example, if a young person is engaged in a training course in the first few months of their participation in the Reboot programme, but is then NEET for the remainder of the evaluation period, they will not have achieved the sustained EET outcomes that the programme aims to support.

Therefore, the outcome indicator(s) should capture employment activities in the latter stage of the participant evaluation period, or education and training activities that are still being completed and are a strong predictor of future employment outcomes, such as the completion of a higher or further education degree. The period captured by the outcome indicator should start late enough into participation that programme activities have time to make an impact, yet early enough to fit into the evaluation period and if possible, to allow for multiple data points related to EET activities to be captured. Our work to explore the relevance and sensitivity of these indicators is summarised at Appendix B.

We analysed the EET activity data of Reboot I participants provided to us by 1625ip to test the relevance and sensitivity of our proposed outcome indicators. This exercise strengthened our confidence that the final 6 months of the first two-year period is an appropriate evaluation window, as we observed important variation in EET status earlier in the participant journey, and that EET status seems to stabilise by the latter stages of this journey.

⁸ 1625ip does feed data on Reboot participants to LAs, thus if doing this, we should be careful not to validate data using its primary source.

3.4 Proposed outcome indicators

Summary: in line with our principles and criteria for indicator selection criteria, we recommended the following outcome indicators for the impact evaluation:

- **EET status (primary outcome):** a binary outcome variable that indicates whether the young person is engaged in part or full time employment or education for the majority of the latter stages of the participant journey.
- **Position on an EET scale (secondary outcome):** a number indicating the level achieved on a predefined EET scale by the young person as a rough guide to the progression away from risk of NEET. The scale will have a fixed number of levels, where each EET outcome is assigned a certain level, with more desirable outcomes attracting a higher score.

These proposed indicators are defined more specifically below, including the data that they would be based on.

Primary outcome variable: EET status

Definition: a binary outcome variable that indicates whether the young person is engaged in part or full time employment or education for the majority of the last 6 months of the two-year evaluation period.

Data used: if this indicator is constructed solely based on LA data, a young person could be considered in EET if they were engaged in an EET activity for at least half of the last 6 months. The period captured by the indicator would thus be between 18 months and 24 months since enrolment in the programme. To collect this outcome there are two possible approaches:

- Defined data collection points every 2 months in the latter stages of the programme for each participant.⁹
- A single data collection point at the very end of the programme that asks participants to detail their EET activity status over the last 6 months.

If this indicator is constructed using HMRC and DfE data we could have access to the start and end dates of the young person's employment and education activities. In this case, the young person could be considered in EET if they were in education or employment for at least half of the days of the evaluation period. However the availability of the DfE data was likely to fall outside the current planned evaluation window.

Strengths: this indicator captures whether the young person obtains and sustains employment or education over multiple months at the end of the participant journey, which are the main outcomes we are interested in. It should be feasible to collect this data for both

⁹ Assuming we cannot control when the first touchpoint is in relation to the young person's start at Reboot, a six month period is required to ensure that there are at least three touchpoints within the capture period. The time between the first and the last touchpoints will be 4 months. (e.g. month 19, 21 and 23)

the Reboot cohort and a comparison group. EET status is also a commonly used metric in other studies¹⁰, and easy to interpret, which would facilitate the communication of findings and comparisons to other studies.

Limitations: the outcome does not capture whether the young person is progressing in employment, or how meaningful their employment is. For example, it cannot differentiate between quality of education or employment, part-time or full-time work, or the reason why the young person might not be in EET. As it is a binary variable, the amount of variation it captures is inherently limited.

Secondary outcome variable: position on EET scale

Definition: a number indicating the level achieved on a predefined EET scale by the young person at the end of the evaluation period. The scale would have a fixed number of levels, and each EET activity would be assigned a certain level, with outcomes that are stronger predictors of long-term success in sustaining meaningful employment being assigned a higher score.

Data used: the scale is adaptable to the data that is available. Thus, the better the data, the more sensitive and relevant the outcome measure can be. If based only on LA data the categories from the EET activity variable in LA data can be used (see Table A2 in the appendix), to construct a scale with the levels depicted in Figure 2.

In response to the need to capture and organise a multitude of EET activities along a single progression scale, any EET scale requires choices on the relative value or meaningfulness of different EET activities. The scale in Figure 2 results from such choices, including:

- Full-time activity is always placed higher than part-time activity, all else being equal.
- Higher education or employment is placed higher than an apprenticeship or other education.
- Progression from any level to another is equally meaningful¹¹.

¹⁰ For example in the [activity agreements pilot](#) by the Department for Education or [this research report](#) on the impact of the Youth Contract provision for 16- and 17- year-olds not in education, employment or training evaluation by the Institute for Employment Studies

¹¹ This original scale was designed in the absence of data on levels of qualification and/or job quality, which are critical to any judgement of relative value for each level of the scale. Since the delivery of the Feasibility Report, a second version of the scale has been agreed with YFF in which it is collapsed into three levels. The outcome is the position on the scale, ranging from 1 to 3, constructed using the same data as the primary outcome:

1: YP is NEET (where NEET/EET status follows the same definition/rules of the primary outcome)

2: YP is in part time EET (see the table below for more info)

3: YP is in full time EET (see the table below for more info)

The option above ensures each level objectively demonstrates more progress toward sustainable employment than the previous level, which required some subjectivity in its interpretation.

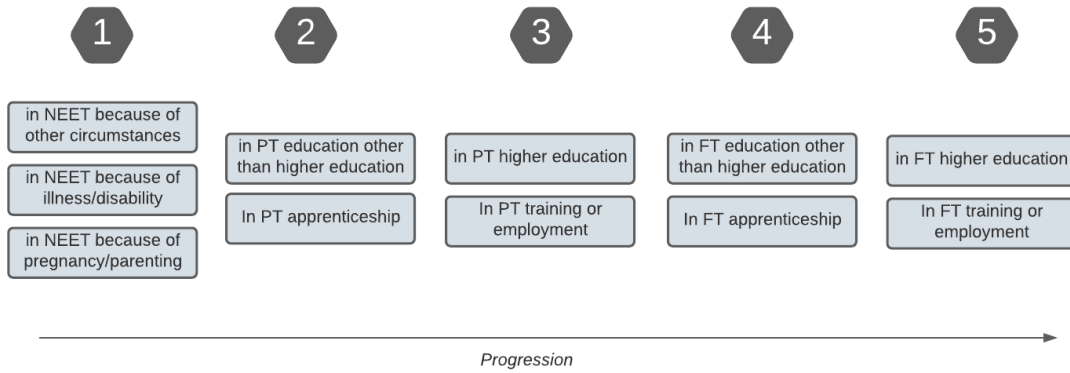


Figure 2. EET scale based on LA data

If based on HMRC data the employment categories could be refined by using the young person’s earnings or hours worked (see Figure 3). For example, earnings could be used to categorise employment based on the number of hours someone earning the national living wage would have to work in order to earn the equivalent amount (this would be similar to indicators used in the SIB targets for Reboot I). The exact hours used in the earnings category are flexible and should be determined based on further research and discussions with 1625ip.

We could similarly refine the education categories if the indicator was based on DfE data. One likely refinement would be to split up the ‘education other than higher education’ category, for example by splitting it up by qualification level.¹² However, at the time of writing the feasibility report, the availability of this data is likely to fall outside the current planned evaluation window.

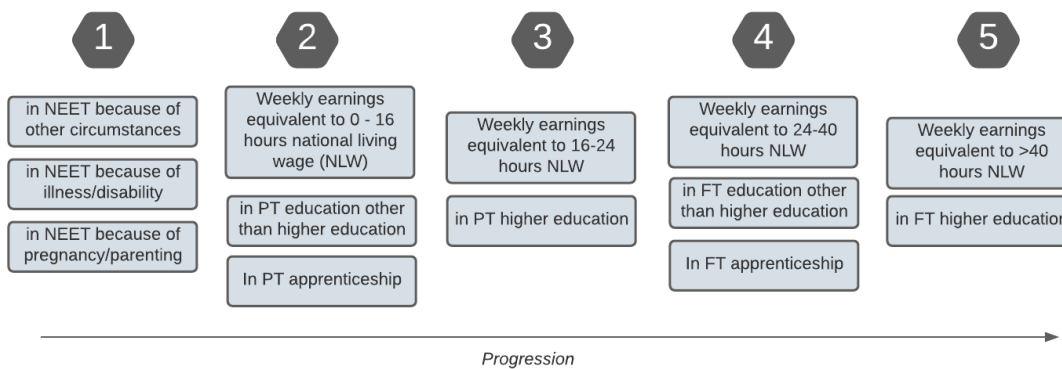


Figure 3. EET scale based on LA and HMRC data.

Justification: the scale allows for more sensitivity in the outcome measure than a binary outcome variable, while still having a single outcome measure that can capture all relevant types of EET activities. It is easily adaptable to the data available and the more datasets we have access to, the more we can refine the scale. It is still a usable outcome measure if we only have access to LA data.

¹² Qualification levels defined by the UK government [here](#)

Limitations: constructing the scale and determining the level each activity corresponds to involves a certain level of subjectivity. The necessary assumption that the distance between each level is equivalent is significant. Additionally, the outcome indicator is not easily interpretable, particularly for audiences not familiar with the Reboot programme or LA data categories, and will likely require thoughtful explanation when describing findings from the impact evaluation.

One way to address these limitations would be to use multiple scales to cover each component of education, employment and training, as this would avoid the subjective judgements involved in determining what 'level' of employment (for example) is comparable to a certain 'level' of education and/or training. These scales would be lower powered to detect significant changes and cannot be interpreted without the context of all outcome variables¹³.

¹³ For example, a decrease in earnings could be due to unemployment (a negative change), or a move into education (a potentially positive change), so the scales would need to be interpreted in relation to other outcomes.

4. Considerations for evaluation design

4.1 Key considerations

Summary: throughout the feasibility study, we aimed to build a shared understanding with the 1625ip team of the trade-offs and requirements associated with the different evaluation design options. To guide these conversations, we focused on four factors:

1. **Sample size**
2. Potential **comparison groups**
3. **Data availability**
4. Implementation **feasibility**

To guide our thinking in identifying and assessing the feasibility of different evaluation design options, we considered four factors:

1. **Sample size:** the number of care-experienced young people that could be enrolled into the programme, and of other young people that could be included in a potential comparison group.
2. **Potential comparison groups:** the possibility of randomising access to the programme (or the timing of the access), or of otherwise constructing a valid comparison group.
3. **Data availability:** the outcome indicators and young people's characteristics for which data exists, and which can be reliably collected for all programme participants and a possible comparison group.
4. **Implementation feasibility:** the possibility of tailoring implementation to enable randomised or quasi-experimental evaluation designs (e.g., randomising certain programme elements; adjusting timelines or participant recruitment; varying programme delivery or intensity across participants).

When considering different evaluation designs for an impact evaluation, there is usually a trade-off between causal power (which is significantly influenced by the first three factors above), and feasibility. Practical or moral barriers can make the implementation of certain designs challenging. This trade-off is illustrated in Figure 4.

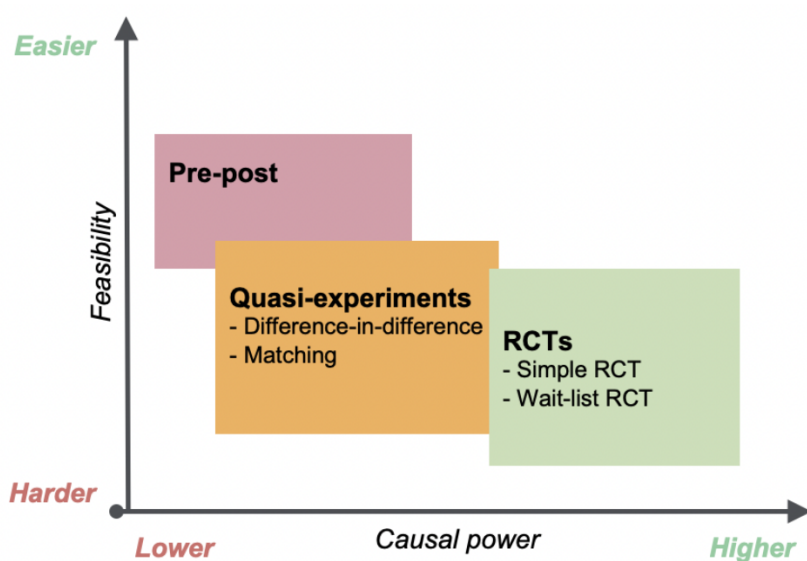


Figure 4. Evaluation design options and associated trade-offs for causal power and feasibility

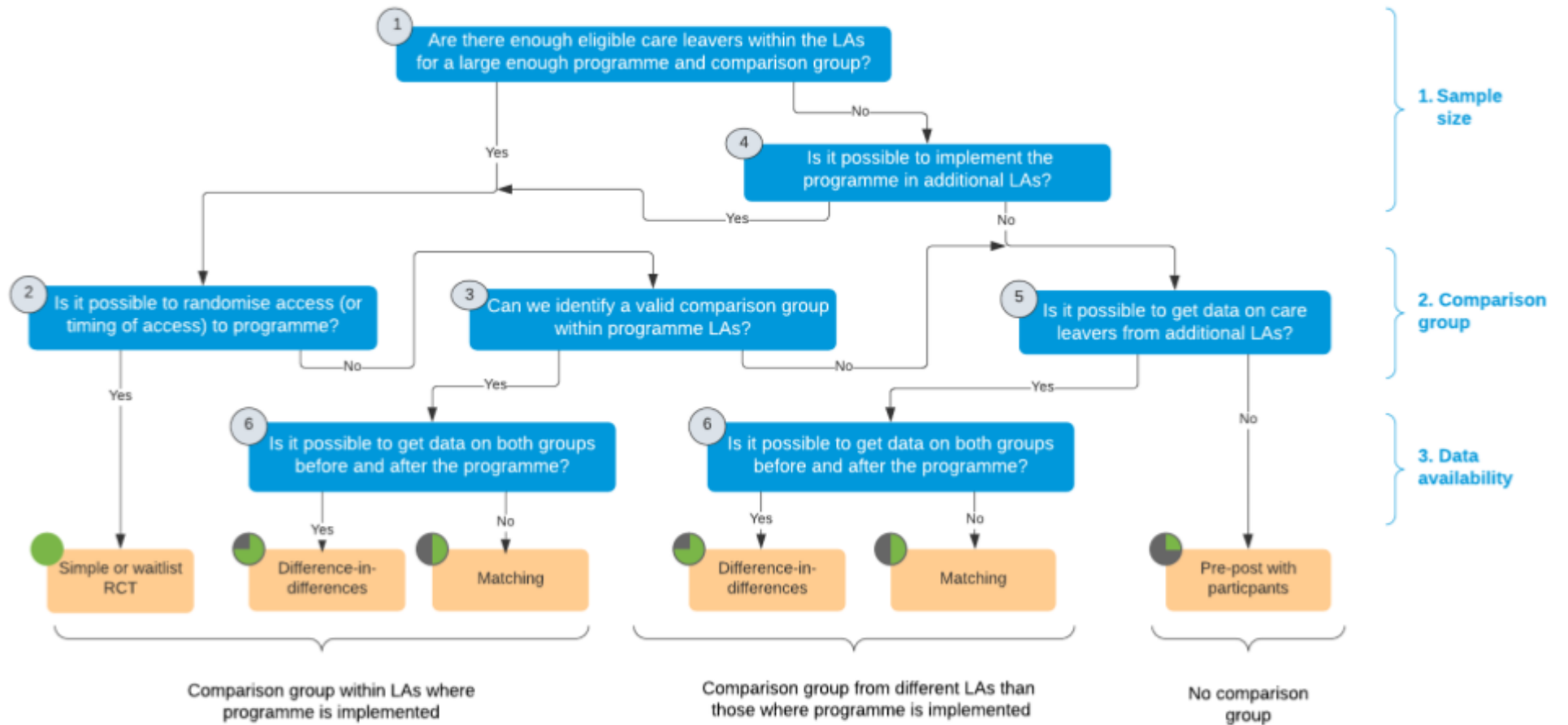
Randomised controlled trials (RCTs) are the most robust evaluation methods to determine causality, however they also have the strictest requirements, including randomising programme enrolment. If randomisation is not feasible or desirable, quasi-experimental designs can be used to estimate causal impact, however they also have important requirements, described in Table 2. Additionally, statistical power suffers under a quasi-experimental design, and, depending on the design, a larger sample would likely be required than under an RCT, all else equal (see Hu and Hoover, 2018). We assessed the feasibility of an RCT and quasi-experimental designs based on the analysis of the factors described above, and the sample size and other requirements associated with each design.

Table 2: Possible quasi-experimental designs and associated requirements

Design	Comparison group requirements	Data requirements
Matching	There exist non-participants who can serve as a comparison group, and reliable data on them exists or can be collected (including prior to programme start)	Outcome indicator data post-intervention and pre-intervention and young person profile data (including many detailed variables correlated to outcomes that can be used for matching), for both groups
Difference-in-differences	Non-participants who can serve as a comparison group (ie. parallel trends assumption can be tested and appears to hold), and reliable data on them exists or can be collected (inc. prior to programme start). Any changes over time that affect the treatment group must equally affect the comparison group.	Outcome indicator data pre- and post-intervention (and pre-intervention young person profile data if possible), for both groups
Regression discontinuity	All sites strictly apply the same eligibility criterion, and this criterion is "arbitrary" (uncorrelated to outcome other than through effect on eligibility); and there are sufficiently 'almost-eligible' young people that we can collect data on before and after the intervention	Outcome indicator data post-intervention and pre-intervention young person profile data (including eligibility criteria), for both groups (eligible and non-eligible)

Pre-post	No comparison group (compare participants before and after programme). It does however require variation in the pre period outcome.	Outcome indicator data pre- and post-intervention (and young person profile data if possible) for participants. Ideally with multiple pre periods. Ideally with multiple pre periods
----------	---	--

Throughout the feasibility study phase, we aimed to build the 1625ip team's understanding of the trade-offs and requirements associated with the different evaluation design options. This was intended to help ensure that the evaluation approach we recommended was tailored to the feasibility constraints and specific evidence needs of the programme, and identify opportunities to tailor programme delivery to facilitate rigorous evaluation. To support discussions with 1625ip, we created the decision tree presented in Figure 5.



: the greener the circle, the higher the causal power

Figure 5. Decision tree to navigate evaluation decision trade-offs

4.2 Sample size

Summary: we conducted power calculations to estimate the smallest effect of the programme on our proposed outcome indicators that an impact evaluation would be able to detect under different assumptions.

- Robust studies of comparable interventions saw effect sizes of **2-13 pp** on EET status/outcomes. Our power calculations suggest that with 250 in treatment and between 144 - 417 participants in control, **we will be powered to detect an MDE between 12.6pp - 9.5pp. This highlights the importance of leveraging all strategies to increase recruitment where possible.**
- The **MDES on the secondary outcome (position on EET scale) ranged from 0.29 and 0.69 points on the 1-5 scale across the scenarios considered.** As a reference, the change observed in this outcome among Reboot I participants over two years (without a comparison group) was 0.63. As this is a highly context-specific outcome indicator, there were no relevant studies to refer to as a comparison point.

These power calculations are based on a number of assumptions. The results point to a risk that the analysis may not be able to detect a positive programme effect even if participants do in fact benefit from the programme, if recruitment targets are not reached.

Two main levers to reduce this risk are i) maximising the number of programme participants and the size of the comparison group, and ii) obtaining data on individual covariates strongly related to the outcomes.

As indicated in Figure 5, an essential first question to understand the feasibility of any evaluation design option is whether there are enough eligible care leavers within the current LAs for a large enough sample made up of a programme and comparison group. To begin addressing this question, we conducted power calculations to estimate the smallest effect of the programme on our proposed outcome indicators that an impact evaluation would be able to detect under different assumptions. The smaller the minimum detectable effect size (MDES), the 'better powered' the evaluation is.

Calculating the MDES requires making assumptions about two key parameters:

- Sample size (number of young people in the programme and the comparison group)
- Average outcome value for the comparison group at 'endline' (latest point in the evaluation period at which outcome data is collected)

We considered different possible scenarios for each of these parameters, as described in Tables 3 and 4 below.

Table 3. Scenarios considered for sample size

Assumed value	Assumptions
153 young people	Pessimistic scenario - No young people aged 16-17 or 21+ are referred, no unaccompanied asylum seeking children (UASC) are referred, 36% of the remaining

	young people (aged 18-21) are assumed to be eligible for Reboot
394 young people	Central scenario - No young people aged 21+ are referred, no UASC are referred. Young people aged 16-17 are referred (as are those aged 18-21), and 50% of these young people are assumed to be eligible for Reboot.
667 young people	Optimistic scenario - The central scenario, in addition to referrals of young people aged 21+ and UASC. 50% of these young people are assumed to be eligible for Reboot.

Note: These values are based on the estimated number of eligible care leavers in the local authorities

Table 4. Scenarios considered for comparison group endline value

Assumed value	Rationale
Primary outcome: EET status	
30%	30% of Reboot I cohort was in EET at baseline (using our definition for being in EET). We assume the comparison group will remain at this level without Reboot support.
Secondary outcome: EET scale level	
Mean: 1.64, SD: 1.21	Average Reboot I value at start of programme (no improvement in control group), standard deviation equivalent to standard deviation of start values
Mean: 2.04, SD: 1.65	Reboot I value at start increased by 0.4 (two-thirds of Reboot I cohort increase), standard deviation equivalent to standard deviation after two years

We calculated the MDES for our proposed primary and secondary outcome indicators for each of these scenarios.

Primary outcome variable: EET status

The MDES under each of the power parameter scenarios is described in Table C1 at Appendix C. The MDES ranges between 10 percentage points (pp) and 21pp across scenarios.

- Our MDES is **12 - 14 pp** in the scenarios using the central estimate of eligible care leavers.
- The proportion of Reboot I participants who would have been considered in EET according to the proposed indicator definition (in EET at least 3 of the last 4 measure points) **increased by 11pp**, from 30% in the first 6 months of Reboot to 41% at the last 6 months of the two year period. However this was not a robust impact estimate as no counterfactual group could be compared against. Note also that Reboot might already have had an impact on the young people's EET status during the first 6 months, and thus this figure might underestimate the true impact of the programme.
- Robust studies looking at the impact of social care programmes on EET outcomes are rare. **The most similar ones we found saw effect sizes of 2-13 pp on EET status/outcomes.**^{14,15,16} Therefore, if the Reboot programme has an effect on EET

¹⁴ Nafilyan, V., Newton, B., Speckesser, S., Maguire, S., Devins, D. and Bickerstaffe, T (2014) *The Youth Contract for 16-17 year olds not in education, employment or training evaluation*. [online] Department for Education.

¹⁵ Alzua, M., Cruces, G. and Lopez-Erazo, C. (2013) *Youth training programs beyond employment*. Mimeo: Evidence from a randomized controlled trial.

¹⁶ Zinn, A.E., and Courtney, M.E. (2017) *Helping foster youth find a job: a random-assignment evaluation of an employment assistance programme for emancipating youth*. Child & Family Social Work, 22, 155-164.

status in line with the upper bound of these studies, the sample size of Reboot III would need to be similar to our central estimate of eligible care leavers to detect a statistically significant impact. It should be noted that the programmes in these studies were not as intensive as the Reboot programme.

Given the above, to ensure the trial is powered, on the basis of findings from similar trials, we need to be able to detect up to a 13pp difference between treatment and control arms. As per the table below, this means:

- We will need to secure covariates to improve our power
- We will need 250 participants in the treatment group
- We will need between 144 - 417 participants in the control group. The more participants we have in this group, the better our power. At the lowest end, 144 participants in control ensures we are powered to detect an MDE of 12.6pp.
- We will need to ensure that outcome data is collected for all participants

Given the importance of recruitment to power, and to ensure a sufficient sample size is reached, we advised:

- Working closely with PAs at LAs to ensure over 21s are recruited, and to maximise recruitment from this group
- Providing LAs with support to collect outcome data from participants, including ensuring regular contact with participants - especially in the control group
- Creating touch points across the onboarding window to monitor recruitment numbers and consider extending the onboarding window as required

Secondary outcome variable: position on EET scale

The MDES under each of the power parameter scenarios is described in Table C2 at Appendix C. The MDES ranges between 0.24 and 0.75 across scenarios. As a reference, the difference between the outcome value for Reboot I participants at the start of the programme and two years later was 0.63, from 1.64 to 2.23.

Given the fact that this outcome measure is tailored to the data available from LAs and the specific aims of the Reboot programme, there are no studies that can be used to estimate what a plausible effect size may be for this indicator. In theory, this outcome measure should be better powered than our chosen primary outcome measure. This is because it not only captures movements between EET and NEET, but also movements within EET (e.g. from part time to full time employment).

As previously mentioned, these power calculations are based on a number of assumptions that can be refined as we understand more about the data available for the analysis. The results point to a risk that the analysis may not be able to detect a positive programme effect even if participants do in fact benefit from the programme. Two main levers to reduce this risk are i) maximising the number of programme participants and the size of the comparison group, and ii) obtaining data on individual covariates strongly related to the outcome indicators.

4.3 Comparison group

Summary: a comparison group is required for all but one of the evaluation design options considered.

- A **valid comparison group can be created through randomisation**, though this has implications for implementation and access to the programme.
- An **alternative is to construct a comparison group through data available on care leavers**. For this to be possible, a group of care leavers of a similar size and with similar characteristics on average to Reboot participants, but who do not receive the programme, must exist within the LAs where the study sample is drawn.

A comparison group is required for all but one of the evaluation design options considered. Meeting this requirement is dependent on being able to identify a **valid** comparison group. In order for this group to be valid, it must be similar on average to participants on characteristics linked to EET outcomes. It must also be sufficiently large to match the number of care-experienced young people participating in the Reboot programme.

A comparison group can be created through randomisation, whereby young people who are recruited are randomly assigned to either the Reboot programme or a comparison group. A comparison group created in this way gives us the highest confidence that the comparison group is similar to the Reboot cohort in both observable and unobservable characteristics, which means that any difference in outcome between the two groups can be realistically attributed to the Reboot programme. In practice, randomisation means that LAs need to refer more eligible young people to the programme than there are places (not counting referred young people who are determined not to meet the criteria), and that up to half of the young people deemed eligible will not be enrolled into the programme.

Having a comparison group that is similar in relevant observable characteristics is required to be able to claim causality in a matching design. However, for a difference-in-difference design, it is strictly only required that the outcome trend (the difference between the pre-programme and post-programme measurement) would be the same in the absence of any intervention. The likelihood of this being true is highest when the programme and comparison group are similar in observable characteristics, thus both of these quasi-experimental designs, need to have a comparison group that is as similar as possible on average as the programme group on all relevant characteristics.

4.4 Data availability

Summary: in addition to being necessary to measure and track our proposed outcome indicators, **data availability is an important determinant of the feasibility of different evaluation design options**.

The RCT approach requires data on EET outcomes to be collected for both those in the treatment and control group. Some of this is already done as part of routine LA data

collection, but additional resource is needed for LAs to ensure that sufficient data is collected.

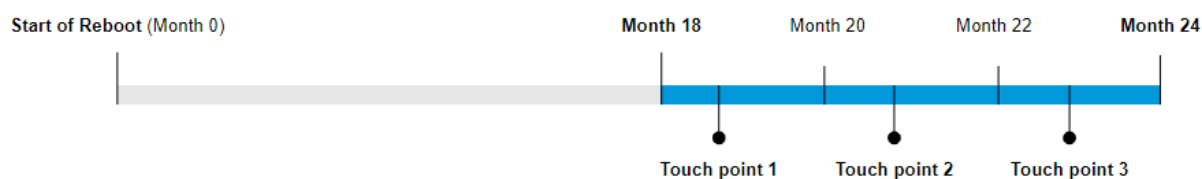
The quasi-experimental designs considered are dependent on the feasibility of using LA data to construct a valid comparison group, and to account for young people characteristics that may be linked to EET outcomes in the analysis. If collected as per statutory requirements and of sufficient quality, the data collected by LAs could satisfy the first purpose, and contribute to satisfying the second purpose. However there are significant risks with relying on data from LAs outside of WECA and North Somerset to construct a comparison group, as would be required in a quasi-experimental design.

In addition to being used to measure and track the outcome indicators proposed in section 3, the data available for the evaluation is important to determine the feasibility of different evaluation design options. Specifically, understanding the data available is important to determine whether it is possible to construct a comparison group, and to account for young people characteristics that may be linked to EET outcomes in the analysis.

Measuring outcomes

EET activity status is recorded each time a personal adviser meets with a young person. This is expected to happen every two months, however our analysis of care leaver data provided by one of the authorities suggests that data collection is less frequent for most care leavers. In order to measure our proposed outcome adequately, updates of EET activity may need to occur more frequently than they do at present, at least in the latter 6 months of the 2 year evaluation window for each participant (treatment and control). We require three touch points during this period. The first one will be between month 18 and 20 of the young person's time in Reboot, the second between month 20 and 22, and the third between month 22 and 24. See the figure below for an example.

Figure 6 Illustrative timing of outcome data collection touch points



Alternatively outcomes could be measured at a single point for each participant at the end of the trial, by collecting self reported EET status over the preceding 6 months. This would however be susceptible to recall error.

In order to measure the secondary outcome adequately we require access to HMRC records. We have reviewed an extract of this data and have confirmed that the dataset contains the variables required for the secondary outcome to be measured.

Constructing a valid comparison group

Using randomisation to construct a valid comparison group carries no additional implications for data collection as long as it is borne in mind that data collection on the control group needs to be as comprehensive as for the treatment group.

The alternative approach of using a comparison group from other LAs carries significant risk. Firstly we would require comparison LAs to adhere to identical data collection requirements to the participating LAs for the full duration of the programme, despite there being no apparent benefit to them. We would require consent from care leavers in these authorities which may be challenging to obtain. If either of these conditions did not hold over the whole duration of the programme, the comparison group would not be usable in the final analysis and there would be no impact evaluation. If we were able to collect all the necessary data, there would remain two further obstacles to constructing a valid comparison group. The first would be to agree that the comparison LAs could not alter their agreed level of support for care leavers beyond a baseline 'business as usual' over the course of the programme. Second, the data may indicate that in fact the selected comparison group is in fact different to the treatment group in ways that could not be accounted for in the analysis¹⁷ - we would not be able to assess this sufficiently until the end of the programme.

With a non-randomised evaluation design, LA data would also be used to ensure that the comparison group constructed using this data is as similar as possible to Reboot participants on observable characteristics plausibly correlated to EET outcomes. The more characteristics data is available for, the greater our ability to maximise similarities between the programme and comparison group. Based on our review of the data held by LAs, characteristics that we would expect to have data on include age, gender, accommodation, missing incidents and police convictions.

Reducing unexplained variation in the outcome

Whether a young person has a positive EET outcome will be determined by many factors other than the effect of participation in Reboot III. Adjusting our analysis to take into account these factors (using 'covariates') can help to reduce unexplained variation in the evaluation outcomes, which can increase the statistical power of the evaluation. This means that a lower number of participants is required to be able to detect a given effect size, assuming all else equal, as we can more confidently isolate the specific effect of the intervention.

We expect that covariates from the LA's, DfE, and to a lesser extent, the HMRC datasets will provide predictive power towards EET outcomes. The key covariates within the DfE data (the National Pupil Database, NPD) include KS2 and KS4 attainment as well as absence rates. Based on prior research modelling predictors of NEET status, we expect that sample sizes could be reduced by at least 20% if we adjust our analysis on the basis of these covariates. This requires consent from those in both the treatment and comparison group to access their records from the NPD at the referral stage.

¹⁷ For example the personal characteristics of care leavers in the comparison LAs may be too dissimilar to the treatment groups, and/or there may be local economic factors that mean that care leavers in comparison LAs may experience better/worse outcomes over the course of the programme.

4.5 Feasibility

Summary: we worked with 1625ip and LAs to assess the feasibility of the requirements linked to each of the factors considered for the evaluation design.

- *Sample size:* we conducted a deep-dive to understand what the maximum sample size could be based on possible adjustments to the referral and enrolment process, and the pool of potentially eligible care leavers in the LAs currently implementing the programme. We concluded that **there is a sufficiently large sample of potential referrals on the basis of current and future care leavers. However this estimate carries significant uncertainty as it is based on assumptions around eligibility that cannot be directly tested.**
- *Comparison group:* randomisation across the four current LAs requires significant changes in implementation, and both buy-in and additional work from LAs and 1625ip. LAs have agreed to the necessary implementation and data requirements for a suitable comparison group to be constructed using randomisation.
- *Accessing required data:* **All four LAs are willing to share their data** to facilitate the tracking of the EET outcomes of both Reboot participants and a potential comparison group, as well as the construction of this comparison group.

Feasibility of reaching a sufficient sample size

Reboot cohort size: we conducted a deep-dive to better understand each stage of the recruitment and referral process, and to identify where in this process any challenges or frictions may have an impact on the scale of the total cohort. The process to onboard young people takes place in four stages. During this process, young people are identified by the LA, referred to the Reboot team, and matched with a coach. At each stage, information is collected to ensure the young person is suitable for the programme, allocated to a suitable coach, and to provide enough background to allow the coach and the rest of the team to understand any challenges and relevant history that would enable them to better provide support.

Through the deep-dive process, we identified six areas that could be streamlined to increase the efficiency of the recruitment process. While these factors are associated with programme on-flow, there may be other opportunities to support an increase in engagement (and reduction in attrition) following a young person's formal entry to the programme.

1. The first stage of this process is **Referral**. This is the point at which the LA team sends the Reboot team the names of care leavers believed to be suitable for the programme. There is substantial variation in the number of young people referred by Personal Advisors (PA; the young person's main point of contact within their LA social care team), who enter the programme. This is driven by a number of things; the extent to which PAs are able to integrate the referral process into their work, the fact that some PAs will have more appropriate young people than others, and PA working

patterns. Due to heavy workloads, some PAs may not have as much capacity to familiarise themselves with the process and as a result are also less likely to refer.

2. During the deep dive, the importance of being responsive was raised as a key factor for maintaining the interest of young people referred, to ensure they reached the end of the onboarding process to become an active participant of Reboot. During the **Allocation** stage of the process, before young people are formally allocated to a coach, key details must be confirmed with their PA. While this process can be completed quickly, in some cases it can take a month or more, during which time young people can lose interest or potentially be referred elsewhere. Shortening this wait time could help reduce the risk of young people abandoning the process when they otherwise could be onboarded.
3. The allocation to coaches is usually made as a monthly exercise, with ad hoc allocations made at times when referrals are urgently needed. While using the monthly case review enables coaches to better plan/manage their workloads, ad-hoc allocations are likely to reduce waiting times for young people who are interested in joining the programme and increase the number who go on to be onboarded.
4. **Coach availability** is a constraint due to the capacity in each coach caseload. Where coaches are at capacity, allocations cannot be made until their existing flow of work slows or is reduced. This is a factor that can create a lag between referrals being made and allocations being completed, and results in longer waiting times for young people, increasing the risk of attrition. This could be mitigated by adding additional capacity and expanding the existing team of coaches, and by slowing the rate of referral. Young people require more flexibility at the beginning of the process whilst relationships are being built, so being unavailable for a new young person can mean they are much harder to engage. It is also important that coaches are available for their young people following the sign up and in the relationship building stage.
5. In the **Case Planning** phase, a meeting must be held between the Reboot team and the young person's PA to complete a case planning document that allows the Reboot team to plan support for the young person. At this stage, securing a meeting with a busy PA can take several weeks and sometimes longer, during which time a young person cannot move to the final stage of the process and formally enter the programme, or meet their allocated coach.
6. During the final **Assessment** stage, the coach meets their young person to complete an initial assessment and collect consent. At this point, following completion of an IA, they are formally counted as a programme participant. In some cases, young people do not attend this initial appointment, which can compound challenges with finding an appointment slot to include their coach and PA. The failure to attend rate is estimated at between 30% and 60%.

We discussed these challenges with 1625ip and reflected on three key takeaways from these findings:

- Most of the challenges identified are linked to delays in the referral process, particularly linked to the capacity of PAs. While the aggregate effect of these delays can be substantial, given the relatively small number of young people referred (and the fact that many of these delays kick in after the initial referrals have been made) our view is that improving the efficiency of the referral process would likely only marginally increase the cohort size.
- As a reference, 1625ip reported that 8 young people who were referred and deemed eligible for Reboot II declined to participate in the programme; some of these may have been more likely to participate had the referral process been shorter or more efficient. However this number suggests that while improving the efficiency of the recruitment process may have other benefits, it is unlikely to result in a large increase in cohort size. A more significant driver of cohort size upstream of this recruitment process may be to increase the share of eligible young people who are referred for the programme.
- The key benefit of streamlining the current recruitment process would be that it could reduce the gap between referral and enrolment, bringing programme start dates for participants closer together. This would be beneficial for evaluation as it would mean that the outcome measurement windows for participants would also be better aligned.

Comparison group size: in addition to increasing the Reboot cohort size, a robust trial requires a valid comparison group of a similar size. We received data from each of the four LAs currently implementing the programme to inform our estimations of the size of relevant groups. Each of the local authorities provided us with data to estimate the total size of the number of eligible young people in their areas for March 2023. The estimates of the number of young people in the relevant groups are in Table 5.

Table 5. Estimates on the number of eligible young people for Reboot in March 2023

LA	Care leavers under 21 [^]	Aged 16 - 17 in care	Care leavers 21+ (open cases)	UASC 16 - 20 year old
Total	363	425	386	156

Note: [^]Figures for care leavers under 21 **exclude** young people who already received Reboot support and unaccompanied asylum seeking children (UASC)

Not all of these young people will be suitable for Reboot. Observable characteristics related to Reboot eligibility include:

- **Age:** Although the programme targets care leavers between the ages of 16 and 25, care leavers under 21 years old are most suitable for the programme, as data is more easily available for them and they are all required to have a PA. However, young people 16 - 17 who are still in care, and care leavers 21+ can also be referred.
- **Prior participation in the Reboot programme:** young people who were part of the Reboot I or Reboot II cohort will not be eligible to be part of the third cohort.
- **EET status:** the programme targets care-experienced young people who are either NEET or at risk of becoming NEET. While LA data do not capture risk of becoming NEET, it does capture EET activity, and Reboot II programme data provides an overview of the EET status distribution across participants at enrolment.
- **Unaccompanied asylum seeking children (UASC):** Although UASC are not ineligible for Reboot, many of them are likely not suitable for Reboot due to related characteristics (e.g. not speaking English or moving out of the county). Often there are other services that specifically target this group.
- **Unsuitability for EET:** Some young people are in a situation where they are not able to pursue EET goals, and thus are not suitable for the Reboot programme. Reasons for this include being in prison, pregnancy/maternity, a disability, or illness.

Based on the figures from Table 5 and the above considerations, we created a pessimistic, central and optimistic estimate for eligible young people for Reboot, based on assumptions around which age groups will be able to participate and the percentage of care leavers who will be or at risk of NEET:

- **Pessimistic:** Only care leavers aged 18 - 20 are referred, no UASC, 36% of care leavers are assumed to be NEET or at risk of NEET.
- **Central:** Care leavers aged 18 - 20 and young people in care aged 16 - 17 (who will become care leavers during their time in Reboot) are referred. Care leavers aged 21+ and UASC are not referred. 50% of care leavers are assumed to be (at risk of) NEET
- **Optimistic:** All care leavers aged <25 with open cases, 16 - 17 year old young people in care, and UASC are referred. 50% of care leavers are assumed to be (at risk of) NEET.

The number of eligible care leavers in each of these scenarios are shown in Table 6. These figures suggest that there are potentially **394 eligible young people** for Reboot III across the four LAs under the central assumptions. In the pessimistic scenario there are 153 eligible care leavers, while in the optimistic scenario there are 667. These estimates are subject to some uncertainty:

- They rely on the programme recruiting from those currently in care (i.e. 16-17 year olds), rather than only those aged 18 and over (which has been the focus of 1625ip's previous recruitment for the programme).
- The % of care leavers eligible for Reboot III is assumed to be 50% in the central scenario. This assumption is based on the age 18-20 NEET rates in the 4 LAs, as well as the percentage of Reboot I participants who were in EET at the start of the programme, but there is considerable uncertainty attached to this assumption and it could be higher or lower than 50% in reality.

Table 6. Number of eligible care leavers in each scenario

LA	Pessimistic estimate	Central estimate	Optimistic estimate
Total	153	394	667

Table 7. EET activity distribution at entry into Reboot II programme as of 14 March 2022

Status	Category	% of cohort	Number of YP	% Total
NEET	EET seeker	43%	56	52%
	Volunteering	1%	1	
	Unable (health)	2%	3	
	Not required (parent)	4%	5	
	Econ inactive	2%	2	
EET	FT work	2%	3	48%
	PT work	14%	18	
	Training	3%	4	
	FT education	4%	5	
	FT education - FE	21%	27	
	FT education - HE	1%	1	
	PT education	3%	4	

Feasibility of constructing a valid comparison group

Randomised design: one possible strategy to create a valid comparison group is through randomisation. Throughout our work to understand the Reboot referral and enrolment process, we have identified the following lessons and implications for the feasibility of randomisation:

1. The Reboot programme referral process is conducted in four stages, during which young people referred by the LA have their background data checked, are assigned to a coach, and have their consent collected before they are formally enrolled into the programme. Given this, randomisation would need to happen at a relatively early stage to prevent a significant additional administrative burden to process young people who will not be participating, while ensuring that referrals are appropriate and young people are indeed eligible. In this case, LAs may need to adapt their procedures with 1625ip to accommodate the randomisation procedure. The quality of referrals from LAs needs to be high and consistent, and 1625ip needs administrative capacity to execute the randomisation, and share the result with LAs.
2. If randomised, a proportion of young people referred to the programme will not receive it at all. Referrers within each of the four LAs need to know that young people they refer may not benefit from the programme, which could make referral conversations with young people difficult, impacting referral rates at the source. In discussion with both 1625ip and LAs, it was noted that guidance needs to be provided to help LA teams have these conversations and make it easier for them to refer people while avoiding disappointment if they are randomised into a comparison group. Framing the programme as an opportunity with limited availability from the outset and providing detailed conversation frameworks early were both discussed as a means of supporting LAs with this part of the process.
3. It was noted in discussion with 1625ip that young people lose interest in participating in the programme when too much time passes between their initial referral conversation, and their contact with a Reboot coach. Therefore programme participation needs to commence soon after randomisation.
4. Young people are referred to the programme in part on the basis of need. Given this, there are ethical implications for any design that involves withholding the programme (if a simple RCT) Given our central estimate, the 250 maximum places for Reboot participants suggests that the programme will not be withheld other than for reasons of limited 1625ip capacity.

These considerations require changes in programme implementation, and both buy-in and additional work from LAs who will be making programme referrals. All four LAs have agreed to support the implementation of these changes.

Quasi-experimental designs: an alternative way to construct a comparison group is through working with LAs (whether currently participating in Reboot or not) in the following conceptual steps:

1. Consider the data for care experienced young people held by LAs, and determine the overall number of young people, and retain data from all young people who are potentially eligible for the programme according to the variables related to Reboot eligibility.
2. Use data collected by 1625ip, to identify young people who were referred for the Reboot programme (both those who ended up enrolling and those who did not) to individuals in the LA data (likely based on name and date of birth, as was previously done for Reboot I).

3. Identify young people in the dataset who were not enrolled into the programme but are as similar as possible on characteristics for which there is data available, and could thus provide a valid comparison group.

Additional steps would be taken to assess the extent to which the identified group of non-participants could provide a valid comparison group.

Feasibility of accessing required data

LAs expressed willingness to share their data to facilitate the tracking of their EET outcomes of both Reboot participants and a potential comparison group, as well as the construction of this comparison group. This requires a data sharing agreement with each LA. Data needs to be pseudonymised, which includes replacing young people's names with a unique ID only known to LAs, and ensuring that their identity cannot be uncovered based on the data included. This requires converting their date of birth into age and other transformations to generalise their data. The data security benefits of removing information have to be balanced against the need to meet the data requirements of the evaluation.

Important considerations to explore are the completeness and quality of the data held by LAs. Given that our proposed outcome indicators rely on multiple data points being available for each young person over the last 6 months of the evaluation period, the availability of reliable data for all young people in the sample is an essential condition for the feasibility of the proposed evaluation approach. Relatedly, we understand that 1625ip provides data on Reboot participants to LAs; it is thus possible that the data quality for Reboot participants is higher than for non-participants, which may introduce bias in the evaluation results. We have reviewed the completeness and quality of the collected data from 2 LAs. The data collected are consistent across LAs and include the variables that are relevant to the outcome measures. However our review found that the recording of EET status is not always as regular as every two months. As a result, in order for our proposed outcome measure to be valid, LAs will need to ensure that EET activity status is updated more frequently than is currently the case, **specifically in the latter 6 months of the 2 year programme window.**

In addition, LAs are not required to collect data on EET activity from age 21 onwards. While our central estimate does not include those 21 and over *at the point of onboarding*, it does include those who will turn 21 during the programme period. For this group, LAs will have to collect EET activity data that they would not normally collect.

Access to data from HMRC and DfE for both a programme and comparison is more uncertain at this stage, but is also not strictly necessary for the outcome indicators proposed. Access to HMRC data was granted for Reboot I participants, and there do not appear to be significant differences in the requirements for HMRC and DfE data. However, the justification for requesting access for a comparison group of young people will be different than just for Reboot participants. BIT and 1625ip are currently agreeing data requirements with HMRC.¹⁸

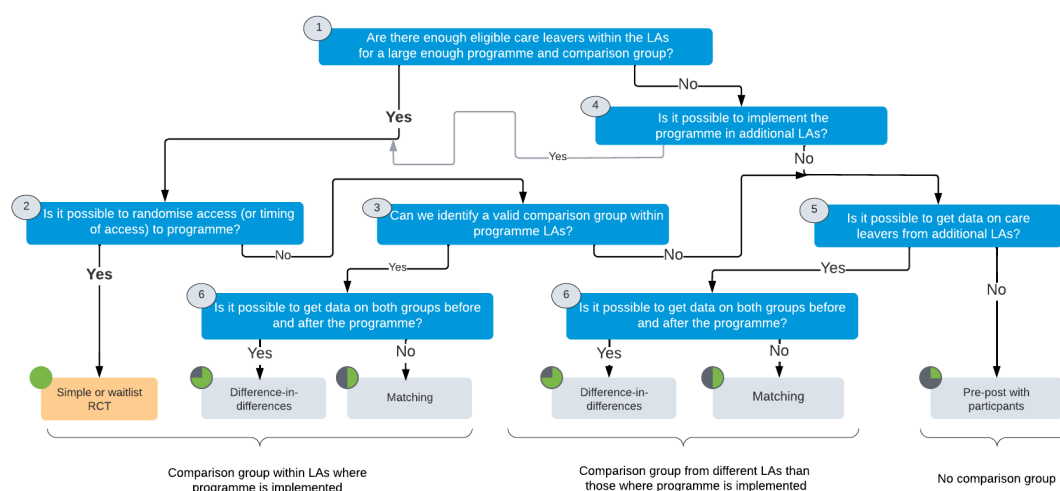
¹⁸ As previously highlighted, access to HMRC data has since been agreed, and forms the basis of our secondary outcome indicators in the design of the full impact evaluation

4.6 Lessons from the feasibility assessment

Summary: an **RCT is feasible**, it does however carry risks: the programme needs to recruit from 16-17 year olds in care and LAs need to ensure that data collection of EET status is expanded beyond its current coverage.

A **matching design** is possible but relies on the cooperation of LAs who would have no benefit from the trial, and carries a major risk that we would only know about the feasibility of the analysis approach once we received the data at the end of the programme.

To outline the lessons from the feasibility study, we return to the questions in the decision tree. While there remain steps to complete to confirm the responses to these questions, some important early lessons emerge.



 : the greener the circle, the higher the causal power. Preferred evaluation designs is in orange

Figure 7: Decision tree to navigate evaluation decision trade-offs

1. Are there enough eligible care leavers within the LAs for a large enough programme and comparison group?

Feasibility: medium Confidence: high

Given the estimated total number of care-experienced young people that could be eligible in the four LAs currently implementing the Reboot programme, and the sample size that would be required to detect effect sizes in line with the optimistic scenarios in our power calculations, we think that a large enough sample of programme participants and a valid comparison group can be constructed within these four LAs, if efforts to recruit participants are maximised, as the mitigation strategies discussed earlier.

2. Is it possible to randomise access (or timing of access) to the programme?

Feasibility: medium Confidence: high

Randomisation poses important challenges in the implementation of the referral and enrolment process for 1625ip and LAs. All parties are aware of these challenges and committed to work through them.

3. Can we identify a valid comparison group within the LAs already involved in the programme?

Feasibility: medium *Confidence: medium*

Based on our analysis of the number of potentially eligible care leavers in the four LAs, the need to maximise cohort size to increase statistical power, and capacity modelling conducted by 1625ip, we believe it is possible to construct a valid comparison group within the current four LAs through randomisation. This requires substantial changes to the referral process for both LAs and 1625ip.

4. Is it possible to implement the Reboot programme in additional LAs?

Feasibility: low *Confidence: low*

Neither 1625ip nor BIT have approached additional LAs about implementing the programme at this stage. As a result, the feasibility of this is unknown, and depends on interest and collaboration from other LAs, the possibility of 1625ip expanding the programme beyond the current four LAs, and additional funding being available to support this. Note that the feasibility of this expansion would be further complicated if it was combined with attempts to randomise access, or timing of access, to the programme (*question 2*).

5. Is it possible to get data on care leavers from additional LAs?

Feasibility: low *Confidence: medium*

While 1625ip or BIT have not approached additional LAs about this at this stage, the level of investment from the four participating LAs, 1625ip and YFF that this would require is significantly lower than expanding the coverage of the programme (*question 4*), as it only requires providing data access, rather than any implementation. It does still however require establishing data sharing arrangements. Accessing data from additional LAs increases the likelihood that a valid comparison group of equal size to the number of Reboot participants in the currently implementing four LAs could be constructed (*question 1 & 3*).

As a reference, in 2021, there were 1175 care leavers between the age of 17 and 21 across the four nearby LAs of Swindon (195), Wiltshire (270), Gloucestershire (394) and Somerset (316). This exceeds the total of 791 care leavers of the same age group in 2021 in the four LAs currently implementing the programme¹⁹. More work would be required to understand the feasibility of constructing a valid comparison group of young people in these LAs (*question 3*), including what support care leavers have access to and broader LA-level differences. This option relies on the participation of LAs who are not currently involved in the programme. For this reason, there is substantial uncertainty as to the feasibility of accessing these data.

6. Is it possible to get data on both the participant and comparison group before and after the evaluation period?

Feasibility: medium *Confidence: medium*

¹⁹ Department for Education (2021) *Children looked after in England including adoptions*. [Database] Available at: <https://explore-education-statistics.service.gov.uk/data-catalogue/children-looked-after-in-england-including-adoptions/2021> [accessed 23/03/2022]

Sharing data at multiple points throughout the evaluation period, including prior to programme start, should not require any additional data collection relative to current activities from LAs, and is thus be feasible if LAs are bought into the programme and willing to provide data. For data from other sources that may be accessed (HMRC, DfE), obtaining data for multiple time points would similarly not require any additional data collection.

5. Recommendations and next steps

5.1 Recommended outcome indicators

Based on our review of available data and our assessment of possible indicators against our criteria of *relevance* to the programme's intended outcomes, the *feasibility* of obtaining data for both programme participants and a potential comparison group, and *sensitivity* (i.e., ability to capture variation across young people and time), we recommended the principles below in selecting outcome indicators for the impact evaluation.

Recommendation 1: select indicators that focus on EET outcomes. We recommended focusing on EET outcomes for the impact evaluation outcome indicators, as they are both highly relevant and much more feasible to obtain data for than non-EET outcomes related to physical and mental wellbeing. However, these non-EET outcomes are relevant to the programme outcomes as well, and should be analysed where data is available to understand the mechanisms of impact and the relationship between EET and non-EET outcomes. To enable this analysis, we are supporting 1625ip to design and pilot non-EET outcome indicators to collect data on for all Reboot participants.

Recommendation 2: select indicators that can be constructed based on data held by LAs, and enriched with data from other sources. LA records are the most feasible source of data, particularly to construct the sample of care leavers for the evaluation. Selecting indicators that can be constructed solely based on this data would help ensure that the data required for an impact evaluation is accessible. However, we also recommend attempting to access HMRC and DfE data to enrich LA data, to obtain covariates that could increase the statistical power of the evaluation. Additionally, we recommend using data collected by 1625ip on all Reboot participants to validate the data obtained from other sources.²⁰

Recommendation 3: select indicators that focus on the latter stages of the programme participant journey. A young person's EET status is variable and is likely to change multiple times over the time of the evaluation period (currently assumed to be two years). In the programme theory of change, participation in EET activities in early stages of the participant journey are anticipated to lead to employment later, and are thus on their own not a signal of meeting programme objectives. Therefore, outcome indicators should, as priority, capture employment activities in the latter stage of the participant evaluation period, or education and training activities that are still being completed and are a strong predictor of future employment outcomes, such as the completion of a higher or further education degree. However, we recommended analysing data collected by 1625ip on EET activities in the early stages of the participant journey to understand participant EET trajectories, and thus identify any lessons for how the programme could further support the achievement of long-term EET outcomes.

²⁰ At time of writing, the approach to the potential inclusion of LEO data had not been agreed. Since the delivery of the feasibility report, the potential for using LEO has increased. This has been reflected in the trial protocol. Should LEO data become available, these data will be used to reconstruct the outcome measure and replace LA outcome data for the analysis of primary outcomes.

In line with our indicator selection criteria and the principles above, we recommended the following outcome indicators for the impact evaluation:






- **EET status (primary outcome):** a binary outcome variable that indicates whether the young person is engaged in part or full time employment or education for the majority of the latter stages of the participant journey (last 6 months).
- **Position on EET scale (secondary outcome):** a number indicating the level achieved on a predefined EET scale by the young person as a rough guide to the progression away from risk of NEET. The scale will have a fixed number of levels, where each EET outcome is assigned a certain level, with more desirable outcomes being scored higher. In addition, due to substitution effects, we recommend separate secondary outcome variables for each of the components of EET to aid interpretation.

We have tested the relevance and sensitivity of these proposed outcome indicators using Reboot I programme data. While these indicators can be based strictly on LA data if it is sufficiently complete and reliable, the feasibility of a rigorous evaluation would likely be significantly higher if HMRC and DfE data are accessed and used to enrich LA data and improve experimental power. However the availability of the DfE data for outcome measurement is likely to fall outside the current planned evaluation window.

5.2 Recommended evaluation design

Based on our assessment of the sample size and data that could be accessed, and taking into account the most feasible approaches to construct a comparison group, we have assessed all possible options for impact evaluation designs. In Table 8 below, we set these out in ranked order by which we understand to be most appropriate given the constraints described in section 4. We discuss our top two design recommendations below, followed by a further discussion of the rationale for the relatively low ranking of the remaining five design options. In addition to the risks flagged for each design, an important risk relevant to all design options is that the quality or completeness of the LA data is insufficient for the envisioned purposes, and that we cannot access suitable data from HMRC and DfE.

Table 8. Assessment of evaluation design options

#	Design option	Causal power	Feasibility	Main feasibility risks
1	Simple RCT in current 4 LAs	High 	Medium	Complexity of implementing randomisation procedures between 1625ip and LAs
2	Matching with comparison group from additional LAs	Low to Medium 	Medium	Low power due to insufficient Reboot cohort size; quality of the LA data and access to DfE and HMRC data; insufficient data to match participants to young people in other LAs
3	Difference-in-difference with comparison group from additional LAs	Medium 	Low	Low power due to insufficient Reboot cohort size; quality of the LA data and access to DfE and HMRC data. Lack of sufficient pre-period data and violation of parallel trends assumption.
4	Simple RCT in current 4 + additional LAs	High 	Low	Complexity of implementing randomisation procedures, service delivery required from scratch in additional LAs
5	Difference-in-difference in current 4 LAs	Medium 	Low	Unlikely to have sufficiently large sample and a valid comparison group
6	Matching in current 4 LAs	Low to Medium 	Low	Unlikely to have sufficiently large sample and a valid comparison group
7	Pre-post in current 4 LAs	Low 	High	Minor relative to other options, however would limit ability to estimate causal impact

Recommended option: Simple RCT within LAs

We believe that the evaluation design that best balances rigour and feasibility is a randomised controlled design where the comparison group would be constructed within the four participating LAs. Given the estimated number of potentially eligible young people, a sufficiently large and valid comparison group of non-participants looks possible within the four LAs currently implementing the programme.

Alternative design: matching or difference in differences, using data from additional LAs

An alternative evaluation design with similar implications for programme implementation is a quasi-experimental design, which draws on data from additional LAs to construct a comparison group. These designs aim to compare individuals that do receive the intervention with individuals that do not but who are otherwise as similar as possible in terms of observable characteristics for which data is available. With these designs, no change to programme implementation would be required.

However, in order for this approach to be viable, we need to access relatively rich data on all individual characteristics relevant to both programme eligibility and the proposed EET

outcome indicators. Based on our review of the data held by LAs, our proposed main source of data, the feasibility of obtaining such rich data on young people's characteristics is uncertain, and would likely require access to other data sources including HMRC or DfE data.

In addition, using a comparison made up of young people outside of the programme creates a risk that differences between LAs may influence the evaluation results. This risk is present for both matching and difference in difference designs.

Though matching allows us to approximate a comparison group that shares characteristics with the programme group, we could not control for wider influences linked to the young person's parent LA. The difference-in-differences design relies on an assumption regarding these influences that can be tested, but there are reasons to think that this design is unlikely to be valid. Firstly, there will be little to no variation in the pre-treatment outcomes, as prior to the programme a large proportion of participants will be of mandatory school age or within the Participation Age where EET rates are high. Secondly, difference in differences may be less valid when the outcome is measured further on in time from the point of intervention, especially as for many of the participants there will be other changes occurring between the intervention and the outcome measurement, such as leaving care. Both these factors present a risk that the before/after comparison used in difference in differences may be contaminated by unobserved differences between LAs that cannot be controlled for in the design.

Given the limitations related to feasibility and causal power relative to a matching design, we assess a difference in differences design to be a less preferable option. In both approaches, we would need to fully understand services available to care leavers in these additional LAs.

Other methods

The discussion above focuses on our three highest ranked options, as additional limitations rule out options 4-6. The remaining option is a pre-post design (option 7), in which no comparison group is required. While this design has high feasibility and relatively minor requirements in terms of implementation, it cannot account for changes over time and thus yields a relatively less rigorous estimate of causal impact.

5.3 Next steps

Following the first version of this report in March 2022, we discussed our recommendations with YFF and 1625ip, as well as the four participating LAs, and all parties have now agreed in principle to support a randomised controlled trial of the Reboot III programme that would commence in mid-2023.

Below we have listed several actions that must be completed ahead of an evaluation:

1. **Designing and testing a new referral process.** For randomisation to take place, both 1625ip and LAs will need to implement new referral procedures, including embedding a randomisation tool and ensuring that PAs are able to have sensitive and productive conversations with young people about the research and the possibility that they may be allocated to a control group. At the time of writing, we are working with 1625ip and all four LAs to implement these changes, and plan to pilot the new

processes with a small subset of young people in early 2023 in order to identify any improvements before a full evaluation takes place.

2. **Establish data sharing agreements between BIT and the four participating LAs.** These will be required for BIT to collect outcome and monitoring data for young people referred to the programme. All four LAs have agreed in principle to establish data sharing agreements and these discussions are ongoing. Data sharing procedures will be tested as part of the planned pilot in early 2023.
3. **Further explore requirements and process to access HMRC and DfE data.** 1625ip and YFF have already submitted a joint request to access HMRC data for Reboot II participants. We will support them to understand the requirements and process to apply for DfE data, and to prepare materials necessary to apply for access to this data. We will continue to support arrangements to access HMRC data for future Reboot III participants.
4. **Obtain ethical approval for the evaluation.** Given the potential ethical concerns about withholding Reboot support from young people allocated to the control group, ethical approval will be required to ensure the evaluation can take place. At the time of writing, BIT are currently seeking ethical approval for the planned pilot through their internal ethics review process. For the full evaluation, YFF and BIT intend to seek approval through the What Works Centre for Children's Social Care review process.

Appendix A. Additional figures and tables

Table A1. Social impact bond targets

Category	Target
Engagement	Enters programme and receives Initial assessment
	Receives assessment and review every 3 months
Education	Begins education / training course
	Completes 25% of course
	Completes 50% of course
	Obtains level 1 qualification
	Obtains level 2 qualification
	Obtains level 3 qualification
	Begins University / Higher Education
	Completes the 1st Year of University / Higher Education
	Completes the 2nd Year of University / Higher Education
	Obtains a Level 4, 5 or 6 qualification including University degree
Employment	Care leaver enters employment
	Care leaver earns the equivalent of 6.5 weeks x National living wage (NLW) x 16hrs
	Care leaver earns an additional amount the equivalent of 6.5 weeks x NLW x 18hrs
	Care leaver earns an additional amount the equivalent of 13 weeks x NLW x 20hrs
	Care leaver earns an additional amount the equivalent of 26 weeks x NLW x 24hrs
Work experience & volunteering	Care leaver enters work experience / volunteering
	Care leaver completes 1 week of work experience / volunteering
	Care leaver completes 4 weeks of work experience / volunteering
Other	Care leaver agrees Education / Employment / Training is right for them
	Care leaver is managing accommodation and related cost effectively
	Care leaver feels safe
	Care leaver has at least one person providing a consistent relationship
	Self-determined outcome

Table A2. EET activity indicator categories used by LAs

Code	Description
0	Not in touch

F1	Young person engaged full time in higher education
F2	Young person engaged full time in education other than higher education
F4	Young person engaged full time in an apprenticeship
F5	Young person engaged full time in training or employment (not including an apprenticeship)
G4	Young person not in education, employment or training because of illness or disability
G5	Young person not in education, employment or training: other circumstances
G6	Young person not in education, employment or training due to pregnancy or parenting
P1	Young person engaged part time in higher education
P2	Young person engaged part time in education other than higher education
P4	Young person engaged part time in an apprenticeship
P5	Young person engaged part time in training or employment (not including an apprenticeship)

Table A3. Indicators collected by LAs on care leavers as per statutory requirement

Item id	Data item name	Description
1CLA	LA number	The LA reference number
2CLA	Motherhood status	Indicates whether or not the looked after girl is a mother
3CLA	DoB mother's child	Date of birth of the first child
6CLA	Gender	Gender of the child
7CLA	DOB	The date of birth of the child
8CLA	Ethnic origin code	The ethnicity of the child
9CLA	Unaccompanied asylum seeking status	Indicates if a child has been an unaccompanied asylum seeking child at any time during the year
11CLA	Child adopted by former foster parents	Indicates whether or not the child is adopted by their former foster carer(s)
14CLA	Legal status of adopters	The legal status of the adopter(s)
19CLA	Reason placement ceased	The reason why the child should no longer be placed for adoption
20CLA	Date of each statutory review	The date of each statutory review
21CLA	Method of participation of each review	The method of participation of each review
22CLA	Previous permanence option	For any child who starts to be looked after from 1 April 2013 onward, have they previously ceased to be looked after due to adoption, a special guardianship or residence order
23CLA	LA where permanence option arranged	The LA code where the previous permanence option was arranged

25CLA	Missing	The status of the child's missing or away from placement without authorisation
26CLA & 27CLA	Missing incident start and end date	The date the missing incident or period of being away from placement without authorisation started and ended
28CLA	Child convicted during year	Indicates whether or not the child has offended in the year ending 31 March
29CLA	Health surveillance checks up-to-date	Indicates whether or not a child's health surveillance or health promotion checks were up-to-date on 31 March
30CLA	Immunisations up-to-date	Indicates whether or not a child's immunisations were up-to-date on 31 March
31CLA	Teeth checked by a dentist	Indicates whether or not a child had their teeth checked by a dentist in the year ending 31 March
32CLA	Annual health assessment	Indicates whether or not a child received their annual health assessment during the year ending 31 March
33CLA	Child identified as having a substance misuse problem	Indicates whether or not a child was identified as having a substance misuse problem during the year ending 31 March
34CLA & 35CLA	Child received / offered an intervention for substance misuse problem	Indicates whether or not a child was offered / received an intervention for their substance misuse problem
36CLA	Strengths and difficulties questionnaire score	This is a single score from the strengths and difficulties questionnaire (SDQ) (only for under 18's)
37CLA	Reason for not submitting strengths and difficulties questionnaire	Reason why it was not possible to submit a strengths and difficulties questionnaire (SDQ) score
38CLA	LA in touch	Indicates if the LA was in touch with the young person i) on, or near, his/her 17th, 18th, 19th, 20th or 21st birthday, or ii) if the young person was in touch with the LA at some point during the year and the child was aged 22, 23, 24 or 25 years during the year
39CLA	Activity	Reflects the young person's main activity status (EET)
40CLA	Accommodation	Type and suitability of accommodation young person is living in
47CLA	Category of need code	The main reason why a child is being provided with children's social care services
48CLA	Home postcode	The postcode of the address where the child was living when they were first taken into care
49CLA	Placement postcode	The postcode of the location where the child is placed
51CLA	Placement type	Indicates, in a broad sense where the child is living
52CLA	Placement provider	Information on the party providing the placement for the child
53CLA	Date episode ceased	The date the episode of care ended
54CLA	Reason episode ceased	The reason why an episode of care ended

55CLA	Reason placement changed	The reason why the placement changed
-------	--------------------------	--------------------------------------

Appendix B. Testing relevance and sensitivity: outcome indicators

We analysed the EET activity data of Reboot I participants provided to us by 1625ip to test the relevance and sensitivity of our proposed outcome indicators.

- Despite caveats linked to the interpretation of missing data, this exercise provided an **indication that the two proposed indicators could adequately capture change in EET activities throughout the programme** across time and young people.
- This exercise also helped to inform our assumptions for power calculations.

The anonymised data received from 1625ip included the start and end date of each recorded EET status for each young person. We then reformatted this data to mimic the format we would expect to receive from LAs: we converted the EET categories used by 1625ip into the categories used by the LAs (see Table A2) and reorganised the data so that it reports the EET status of the participant every two months.

There was a significant amount of missing data, especially in the first few months (when the coach may not have yet entered EET activity information in the database) and after two years since enrolment, when many young people's participation in the programme (and thus data collection) ended. Figure B1 shows the number of young people for whom an activity status was recorded at each 2-month interval (total N = 250)²¹. When an activity status was missing, we recorded the most likely activity status based on the information available. After discussions with 1625ip, we decided to assume that a) if a status is unknown or missing at any certain time, the young person's EET status has remained unchanged since the last known status, and b) if there is no last known status, we assume the young person is NEET.

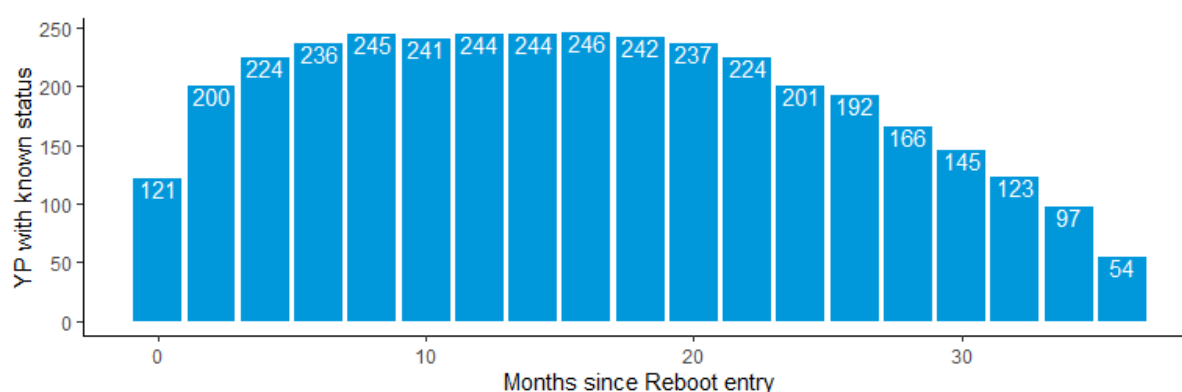


Figure B1. Number of young people with a known EET status at each 2-month measure point

²¹ 1625ip reported that 237 young people completed the initial assessment, and thus had a reported EET status for at least one time point, however the number of observations with at least one known EET status in the data we received was 250. We are awaiting further clarification on the discrepancy, but do not expect that this discrepancy would meaningfully change the conclusions from our review of the Reboot I data.

Primary outcome indicator

We found that in Reboot I there was a sharp increase in EET status in the first year of the programme, after which the proportion of young people in EET stabilised (see Figure B2). 1625ip reported that this did not completely match their expectations, and it is possible that the missing data (particularly at the start of the programme) may bias these observed patterns. In addition, it should be noted that the Reboot I cohort coincided with the Covid-19 pandemic, which likely will have influenced employment and education opportunities.

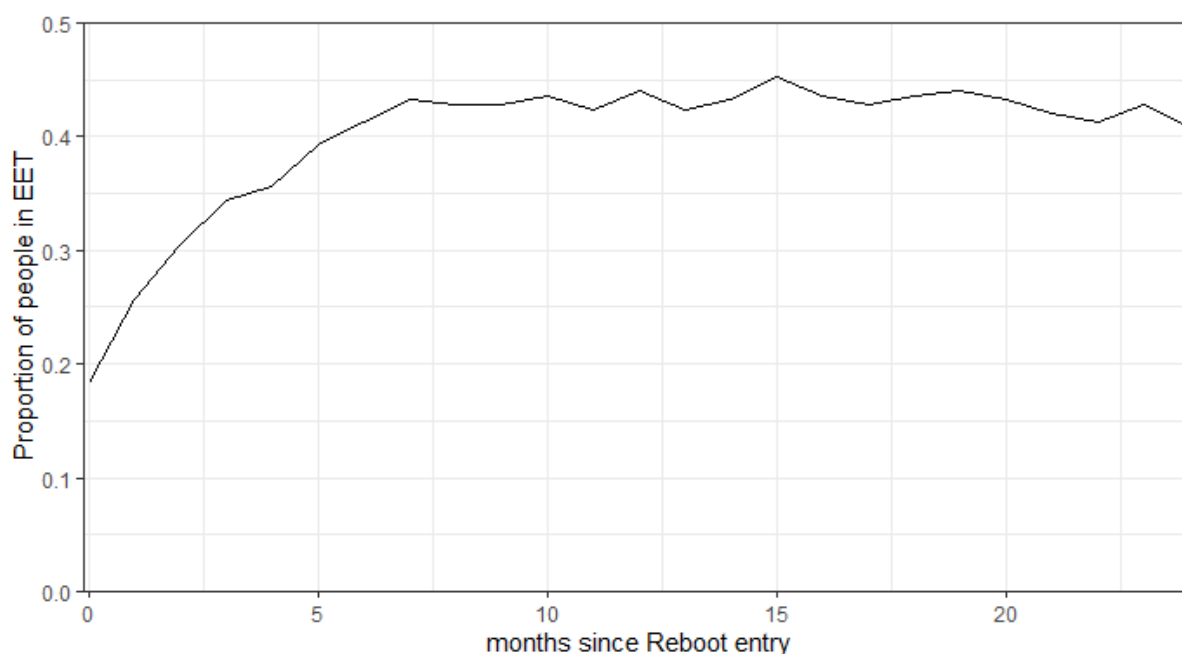


Figure B2. Proportion of Reboot I cohort in EET during the programme (N = 250)²²

Despite these caveats, this exercise does strengthen our confidence that the final 6 months of the first two-year period is an appropriate evaluation window, as we observe important variation in EET status earlier in the participant journey, and that EET status seems to stabilise by the latter stages of this journey.

More specifically, if we apply the proposed definition of the primary outcome to the Reboot I data, we find that 41% of the participants are in EET for 2 out of the last 3 data points. If we instead use 2 of the last 4 data points (and thus increase the measuring period by 2 months) the outcome is 45%. This is shown in Figure B3.

We believe either definition is workable. Using 4 touch points would increase the EET % for both the Reboot and comparison group, but would also increase the measurement period. In the power calculations below we have assumed the outcome is based on using 3 data points over 6 months, illustrative practical timings of the data collection are displayed in figure B4.

²² In this figure and Figure 6, as previously explained, after discussions with 1625ip, we decided to assume that a) if a status is unknown or missing at any certain time, the young person's EET status has remained unchanged since the last known status, and b) if there is no last known status, we assume the young person is in NEET

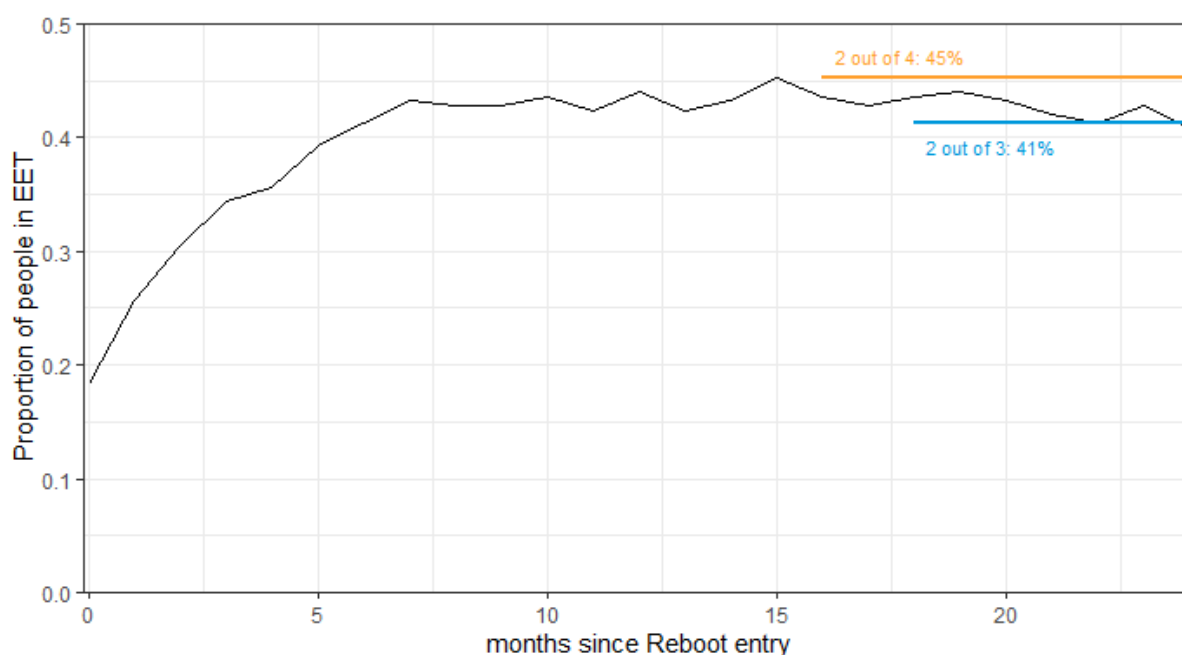


Figure B3. Proportion of Reboot I cohort in EET, with two possible definitions of the outcome measure indicated (N = 250)

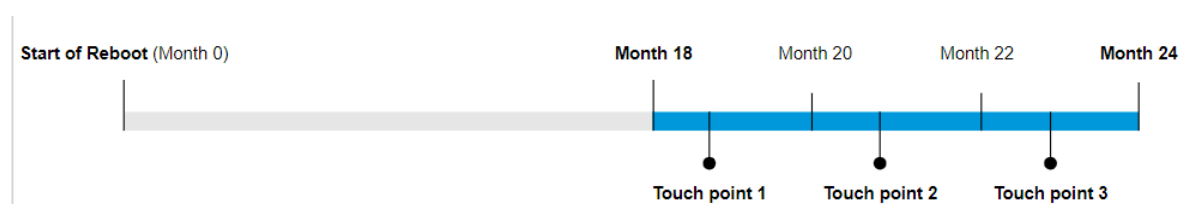


Figure B4. Illustrative timing of outcome data collection touch points

As per Figure B4 above, participants will be contacted at three points during the measurement window. They will be asked whether they are currently in EET, and a binary outcome will be recorded (yes/no). If the response for two out of three of these points is 'yes', this participant will be counted as a positive outcome.

The timing of the measurement points for each participant will be varied, and determined by when each LA contacts each participant to collect data. LAs will be advised to contact participants as close to the start of month 19, 21, and 23 as possible. However, all participants will be contacted three times during the last 6 months of their participation. Contacting participants three times, alongside the provision of incentives for each response, increases the likelihood that we will capture accurate data on EET participation.

Secondary outcome indicator

We also constructed the EET scale, as defined using LA EET activity categories, using Reboot I programme data. Figure B5 shows the distribution was in month 1 and 24 of the programme, and how young people transitioned between the two timepoints.

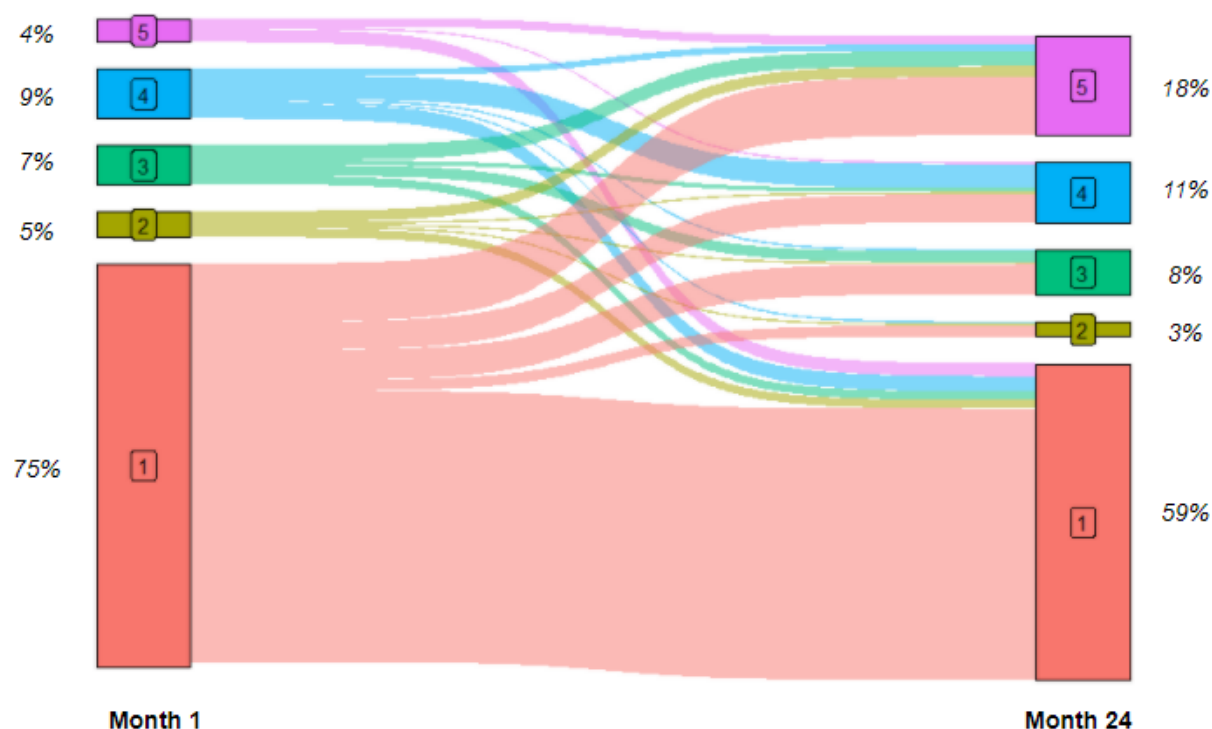


Figure B5. Distribution of EET scale level at start and two years into the programme, for Reboot I participants.

The average level 24 months after enrolment was 2.3 (out of 5), which was 0.6 higher than the average value at month 1. It is clear that most of the change is driven by young people moving from stage 1 to stage 5. There is reasonable variation in the distribution of values across time within the other levels of the scale (2-4).

Similarly to the primary outcome measure, we observe meaningful variation in young people’s position on the EET scale, thus increasing our confidence in the relevance and sensitivity of this proposed secondary outcome indicator.

Appendix C. Power calculations

Results of power calculations for the secondary and primary outcomes as below. Our assumptions for these power calculations are important to note:

- The approach to power calculations varies depending on the evaluation design. The MDES of an evaluation depends on the evaluation method. The calculations in this section are conducted assuming a simple RCT design, as these are both simpler to conduct and interpret, and they are intended to inform a recommendation about the feasibility of conducting an RCT. The logic is similar for the quasi-experimental designs also considered in this feasibility study, matching and difference-in-differences, however the result of the calculation for these designs will differ from those presented below. Specifically, while the MDES for a matching design is likely to be similar to an RCT, it will always be higher for a difference-in-differences design (see Hu and Hoover, 2018).
- In some of our calculations we included predictive power from covariates that correlate with the outcome. Although some of these covariates can be found in LA data, we will need access to DfE data on participants' characteristics to reach the assumed level of predictive power in these scenarios.
- We have not made an explicit assumption on the fraction of the sample we will not be able to obtain endline outcome data for ("attrition"), and thus is 0% in our calculations.
- We assume that the maximum number of Reboot participants is 250, based on YFF's funding intention and 1625ip's onboarding modelling.

Table C1: Power calculation results for primary outcome

Outcome indicator	Predictive power covariates (R^2 value)	# of Reboot participants	# in comparison group	Total sample size	Cohens' H Effect size	MDES	% EET in programme group at endline [^]
Proportion of young people in EET after 2 years since programme start	0	77	76	153	0.45	22.1pp	52%
		197	197	394	0.28	13.6pp	44%
		250	144	394	0.29	14.1pp	44%
		250	250	500	0.25	12.0pp	42%
		250	417	667	0.22	10.7pp	41%
	0.2	77	76	153	0.41	19.7pp	50%
		197	197	394	0.25	12.1pp	42%
		250	144	394	0.26	12.6pp	43%
		250	250	500	0.22	10.7pp	41%

		250	417	667	0.20	9.5pp	40%
--	--	-----	-----	-----	------	-------	-----

Assumptions: using academic standards of power = 0.8 and significance level = 0.05; ^Assuming a counterfactual outcome of 30% in EET in the comparison group.

Table C2. Power calculation results for secondary outcome

Outcome indicator	Average in comparison group	# of Reboot participants	# in comparison group	Total sample size	MDES	Average in programme at endline
Position on EET scale after programme start	Mean: 1.64 SD: 1.21 R²: 0.2	77	76	153	0.49	2.13
		197	197	394	0.31	1.95
		250	144	394	0.30	1.94
		250	417	667	0.24	1.88
	Mean: 2.04 SD: 1.65 R²: 0	77	76	153	0.75	2.79
		197	197	394	0.47	2.51
		250	144	394	0.45	2.49
		250	417	667	0.37	2.41

Assumptions: using academic standards of power = 0.8 and significance level = 0.05