

DFN Project SEARCH: Feasibility Study Report



Table of Contents

About this Report	5
Acknowledgements	6
Executive summary	8
The DFN Project SEARCH programme	8
The feasibility study	9
Findings	10
Programme recruitment and set up	10
Impact evaluation feasibility	10
Recommendations and next steps	12
RCT mobilisation phase	12
DiD data collection	13
1. Introduction	14
1.1. Background	14
1.1.1. Young people with Special Educational Needs and Disabilities	14
1.1.2. Transitioning into work: challenges for young people with SEND	15
1.1.3. Supported internships	16
1.2. The DFNPS programme and selection	18
1.2.1. Programme structure and set up	18
1.2.2. Participant referrals and eligibility	21
1.2.3. Site assessment days	21
Site assessment day tasks	22
1.2.4. DFNPS supported internships	23
1.3. The feasibility study	24
1.3.1. Research questions	24
1.3.2. Report structure	26
1.3.3. Timeline	28
2. Methods	30
2.1. Theory of Change	31
2.2. Data collection and analysis	33
2.2.1. Quantitative data	33
Merged data	35
Quantitative data analysis	36
2.2.2. Qualitative data	36

Observations	37
Interviews with site assessors.....	38
Qualitative data analysis.....	38
2.3. Feasibility assessment.....	39
2.4. Key methodological limitations	40
2.4.1. Triangulation of data sources and missing data	40
2.4.2. Rubric sheet design and rubric data collection	41
2.4.3. Interviewees self-selection bias	42
2.4.1. Lack of research on supported internships themselves	42
3. Programme recruitment and set up	43
3.1. Site assessment days	43
3.1.1. Venue set up	43
3.1.2. Attendance levels	44
3.1.1. Survey completion and consent rates	45
3.1.2. Assessment of candidates	46
Use of the rubric sheet	46
Site assessors' discussion	47
3.2. Candidates' characteristics and outcomes	48
3.2.1. Sociodemographic characteristics	48
3.2.2. Rubric sheet scores and supported internship offers.....	50
4. Feasibility assessment	53
4.1. Outcomes and data sources	54
4.2. Randomised Controlled Trials: general considerations	55
4.2.1. Methodology	56
4.2.2. Enablers.....	57
4.2.3. Barriers	59
4.3. RCT with scale up to participants without EHCPs	61
4.3.1. Practical feasibility and robustness.....	61
Enablers and barriers	63
Sample size	67
4.3.2. Acceptability	70
4.3.3. Working relationships	70
4.3.4. Ethics.....	72
4.3.5. Feasibility assessment summary.....	73
4.4. RCT with scale up to new colleges	73
4.4.1. Practical feasibility and robustness.....	74
Enablers and barriers	74

Sample size	75
4.4.2. Acceptability	75
4.4.3. Working relationships	76
4.4.4. Ethics	76
4.4.5. Feasibility assessment summary	76
4.5. Matched Difference-in-Differences at college level	77
4.5.1. Practical feasibility and robustness	77
Methodology	77
Enablers	79
Barriers	80
Sample size	83
4.5.2. Acceptability and working relationships	85
4.5.3. Ethics	86
4.5.4. Feasibility assessment summary	86
4.6. Regression Discontinuity Design	87
4.6.1. Practical feasibility and robustness	87
Methodology	87
Enablers	89
Barriers	91
Sample size	95
4.6.2. Acceptability	97
4.6.3. Working relationships	98
4.6.4. Ethics	98
4.6.5. Feasibility assessment summary	99
5. Conclusions	101
5.1. Programme recruitment and set up	101
5.2. Feasibility assessment findings	102
6. References	104
7. About the authors	109
Appendix A: Rubric Sheet	110
Appendix B: Survey Script	115
Appendix C: Causal Mechanisms	120
Young people in supported internships	120
Families and carers	121
Employers and host businesses	122

Educators and colleges	122
Local authorities, the State and wider society	123
Key assumptions and risks	123
References	124
Appendix D: Observation guide	126
Appendix E: Feasibility Assessment	130
Robustness of the proposed design	130
Practical feasibility	131
Acceptability of the design	132
Working relationship with the delivery organisation	133
Ethical considerations	134
Appendix F: Other designs considered	135
Randomised Controlled Trial at college level	135
Matched comparator analysis	135
Individual-level Difference-in-Differences	136
Appendix G: Options for outcome data access	138
Appendix H: Economic analysis	140
Appendix I: Supplementary Statistical Analysis	143

About this Report

Youth Futures Foundation is the What Works Centre for youth employment. The organisation's aim is to narrow employment gaps by identifying what works and why, investing in evidence generation and innovation, and igniting a movement for change.

This report presents the results of a study exploring the feasibility of conducting a full-scale impact evaluation of a supported internship programme run by DFN Project SEARCH (DFNPS). DFNPS is a national charity that offers supported internships to young people with a learning disability, who are autistic, or both, to support them secure meaningful, permanent employment.

Youth Futures commissioned researchers at the Policy Institute, King's College London (King's), to undertake the research. The feasibility study ran from November 2024 to October 2025, and a recommendations paper with early insights was produced in August 2025.

This report is structured as follows:

- **Chapter 1:** presents the context and background of the DFN Project SEARCH programme, outlines the key features of the programme, and presents the feasibility study aims and research questions.
- **Chapter 2:** outlines the methodology used to conduct this feasibility study, presents the programme's Theory of Change, and the data collection methods and criteria used in the feasibility assessment.
- **Chapter 3:** provides information on the programme's recruitment and presents the key demographics of the feasibility study participants.
- **Chapter 4:** presents the feasibility assessment of four different evaluation designs that were considered.
- **Chapter 5:** outlines the key conclusions.
- **Chapter 6:** presents the list of references.
- **Chapter 7:** presents the authors of this research.

Acknowledgements

We would like to extend our sincere thanks to the entire DFN Project SEARCH team, especially Claire King and Amy Mitchell, whose unwavering support, motivation and responsiveness have been crucial throughout this project.

We further wish to acknowledge Toslima Khatun and George Kinkead at King's College London for their contributions to the research and support.

We also wish to express our gratitude to the Youth Futures Foundation team, who have commissioned this work, and especially to Dan Muir and Matthew Armstrong, for their guidance, support and excitement about the project. Their commitment to the success of the project has been invaluable for the completion of this study.

We would also like to take this opportunity to remember Dr Jane Colechin (formerly the Deputy Director of Evidence and Evaluation), whose passion for this work continues to be an inspiration.

Contact:

Get in touch

Youth Futures Foundation

Fivefields,
8 - 10 Grosvenor Gardens,
London
SW1W 0DH

Email: evaluation@youthfuturesfoundation.org and
dan.muir@youthfuturesfoundation.org

www.youthfuturesfoundation.org

The Policy Institute, King's College London

Melbourne House,
44-46 Aldwych London,
London
WC2B 4LL

Email: edit@kcl.ac.uk

<https://www.kcl.ac.uk/policy-institute>



Executive summary

This report presents the results of a feasibility study of DFN Project SEARCH (DFNPS), a supported internship programme, and discusses the feasibility of conducting a full-scale, robust impact evaluation of the intervention.

In 2024, researchers from the Policy Institute at King's College London (King's) were awarded a grant from Youth Futures Foundation to conduct a feasibility study centred on identifying and assessing approaches to evaluate the impact of the DFNPS programme in the UK.

The study produced an internal Recommendations Paper in August 2025, which set out agreed criteria for assessing potential impact evaluation approaches and shared initial recommendations on feasible approaches for a full-scale DFNPS evaluation. This final feasibility report elaborates on these initial recommendations and presents the full results of the research, based on subsequent conversations with Youth Futures and DFNPS, and further data analysis.

Overall, the feasibility study shows that **DFNPS is suitable for a full-scale, robust impact evaluation**. While some refinements to data collection and delivery processes are needed, the programme's established infrastructure and scale of delivery, strong partnerships with programme partners, plausible impact on employment outcomes, and clear commitment to the evaluation provide a strong foundation for evaluating its impact at scale, including through an experimental design.

The DFN Project SEARCH programme

DFNPS is a supported internship programme delivered by a charity of the same name. DFNPS works in partnership with employers across the private and public sectors, such as Asda, Amazon and NHS Trusts. The programme consists of supported internships offered to young people aged 16-24 with a learning disability, who are autistic, or both. The overarching goal of the programme is to equip participants with the skills, confidence and experience needed to transition into meaningful permanent employment, paid at competitive wages.

Currently, DFNPS participants must have an Education, Health and Care Plan (EHCP) to be eligible for a DFNPS supported internship and to receive funding for the placement. Typically, participants are identified and referred to the programme by further education (FE) colleges, although Local Authorities (LAs) can also refer eligible participants. Following referral, potential participants attend a site assessment day

involving task-based activities and interviews with a selection team. The selection team comprises representatives from the FE college and the employer hosting the supported internships. Activities and interviews are scored using a structured rubric sheet, allowing the selection team to analyse the profile of each potential participant and to determine whether they are ready for a supported internship.

If candidates are offered the supported internship, they are allocated to a host business where they undertake three placement rotations within different departments of the host business. They also receive on-the-job learning opportunities and support from tutors, job coaches, and the host business's managers during their placement. Supported internships last nine months and start in September each year.

The feasibility study

The feasibility study ran between November 2024 and December 2025, focusing specifically on the DFNPS site assessment days that ran between March and July 2025 to select the 2025 programme cohort (i.e., starting their internship in September 2025). The feasibility study included the following research activities:

- **Desk research** to review existing evidence on supported internship programmes, and the youth employment experiences for young people with a learning disability, with autism, or both.
- **Collection and analysis of quantitative data** on candidates' characteristics through a survey (n=406), attendance at site assessment days captured through monitoring data (n=349), rubric sheet scores (n=399), and data on programme offers and acceptances (n=179).
- **Collection and analysis of qualitative data** from semi-structured interviews with site assessors (n=19) and from observations of site assessment days (n=7).
- **Stakeholder engagement activities**, including three workshops with DFNPS (a Theory of Change workshop, a data mapping workshop, and an impact workshop) plus on-going conversations to gauge acceptability for different impact evaluation designs.
- **Assessment of the feasibility of multiple impact evaluation designs**, ranging from Randomised Controlled Trials (RCTs) to quasi-experimental designs (QEDs), including Regression Discontinuity Designs (RDD) and Difference-in-Differences (DiD) approaches, with the aim of identifying a feasible impact evaluation methodology that could be implemented and concluded by 2028 – in line with Youth Futures' funding requirements – and the set up for a potential pilot impact evaluation.

Findings

Programme recruitment and set up

Observations of site assessment days show that, overall, these were well-structured and designed to assess candidates' workplace skills and programme suitability. Most assessment days provided candidates with tasks mimicking the ones they would have to do in their workplace and used effective methods such as the scoring rubric sheet and assessors' discussions to determine suitability for supported internships.

Both observations and interviews with site assessors highlighted considerable variation across site assessment days related to format, delivery, and available facilities. Notably, we found that there were differences in space and privacy available for candidates to complete their assigned tasks, which can affect their comfort and performance.

Observations also showed that the scoring rubric sheet was valuable for ensuring structure and transparency, and in supporting communication with families about internship offers. However, researchers observed some inconsistency in how the rubric sheet was interpreted by assessors across sites, leading to variability in scoring approaches.

Finally, site assessors reported that assessors' follow-up discussions fostered shared ownership among colleges and employers, and a balanced selection process, enabling assessors to consider candidates' score alongside other information, such as personal circumstances and suitability for specific supported internship opportunities.

Overall, while the research identifies potential to improve consistency across site assessment days, the current candidate selection process – including the site assessment days, scoring through rubric sheets, and assessors' discussions – appears broadly robust and largely can support a robust impact evaluation. Importantly, **improving data collection approaches and internship allocation systems would be essential enablers** for delivering a robust impact evaluation of DFNPS.

Impact evaluation feasibility

Drawing on the analysis of programme delivery and a series of workshops with DFNPS and Youth Futures staff, the feasibility study assessed potential approaches for robustly evaluating the impact of DFNPS. The assessment focused on identifying designs that were practicable and acceptable to DFNPS partners, ethically sound, and methodologically robust. Four designs were explored in particular depth as they were considered the most promising. These are the main focus of this report, while other

designs, which were assessed initially but ultimately deemed infeasible, are covered in the appendix.

First, two different designs for **Randomised Controlled Trials** (RCTs) were explored. These would involve YOUTH FUTURES-funded delivery for the 2026 cohort onwards that (1) scaled up provision to young people with EHCPs in new colleges currently unaffiliated with DFNPS, or (2) scaled up provision to young people with SEND but without EHCPs in colleges with existing partnerships with DFNPS. Eligible candidates would be randomised into a treatment group, receiving DFNPS, or a control group, receiving business-as-usual support from their colleges. The employment outcomes of participants in the two groups would be compared to causally estimate the impact of DFNPS. Overall, the study found that **both RCT approaches are feasible**, helped by the large potential target cohorts and the programme's established infrastructure. However, DFNPS's existing partnerships with colleges and employers means that the scale-up to a new cohort of SEND learners without EHCPs is more feasible in terms of achieving the required sample sizes in a timely fashion.

Second, a **college-level matching difference-in-differences** (DiD) approach – using historical college-level data to compare aggregate employment outcomes between DFNPS-participating colleges and other comparable, non-participating colleges – was also found to be feasible and acceptable. While much less resource-intensive, this design is quite a blunt instrument analytically, simply assessing the extent to which a college implementing DFNPS sees a change in their aggregate employment outcomes for the relevant cohort compared to a comparator group of colleges, while ignoring participant and college context, the rubric score, and work-readiness. The estimates of DFNPS impact would also require stronger assumptions to be met to be considered sufficiently robust. Furthermore, due to limits in the availability of existing data, the design would likely only be able to detect the impact of DFNPS if its magnitude is sizeable, which available evidence suggests it might be, while an RCT could be designed to be powered to detect smaller effects. Despite these limitations, the study recommends **implementing a college-level DiD to complement the RCT**.

Finally, a **regression discontinuity design** (RDD) was also explored. Currently, there is no pre-established threshold at which DFNPS allocate supported internships. Therefore, implementing this methodology would require adjusting the selection process so that DFNPS is delivered to participating cohorts based on whether they meet an agreed rubric score threshold during site assessment days, and comparing the employment outcomes of those who did and didn't receive DFNPS to identify the impact of the programme. Assessing whether an implicit threshold has been applied historically was also in scope. However, subsequent analysis of candidates' rubric scores during the feasibility study identified substantial challenges with small sample sizes around the threshold and risks of sorting (i.e. of scores being manipulated upwards or downwards to

unduly affect the chances that a candidate is offered an internship). This approach was explored in depth due to being potentially valuable as a pilot evaluation design, but ultimately, **an RDD is not recommended for a full impact evaluation.**

Recommendations and next steps

Overall, the research shows that DFNPS is well positioned for a **full-scale, robust, causal impact evaluation, combining RCT and DiD approaches.** Such an evaluation would not only provide high-quality evidence on DFNPS's impact on young people's employment outcomes, but also contribute valuable insights for the wider field of youth employment support. Specifically, we recommend progressing to an RCT mobilisation phase – looking to deliver a trial through scaling-up DFNPS provision to young people with SEND without EHCPs in colleges with existing partnerships with DFNPS – and supplementing the trial with a DiD approach. The option for a scale-up RCT to new colleges also remains feasible, but conversations with DFNPS and YOUTH FUTURES highlighted that the resource burden of the approach would be excessive, especially since the full impact evaluation would have to conclude by the end of 2028 according to Youth Futures funding timelines, limiting the time that could be spent on recruiting new colleges.

The RCT-DiD combination would provide a credible, acceptable, feasible, ethical, robust, and risk-mitigated approach to assessing the causal impact of DFNPS on the employment outcomes of young people with SEND (specifically with a learning disability, with autism, or both). The designs, combined, can deliver more robust evidence, investigating both historical and new DFNPS data, and allow for better phasing in the publication of findings. In addition, we recommend incorporating an Implementation and Process evaluation (IPE) to explore the 'how' and 'why' behind the evaluations of the programme's impact.

RCT mobilisation phase

The mobilisation phase should be used to design key RCT processes with DFNPS. This would include agreeing timelines, recruiting colleges and host businesses, refining eligibility criteria definitions, finalising randomisation and data collection methods, and establishing data sharing agreements. The mobilisation phase would also test the practical feasibility of implementation, including the capacity of colleges to recruit SEND young people without EHCPs, as well as working to increase the validity and robustness of participant surveys through cognitive testing. In addition, several process improvements could be made during the mobilisation phase to deliver a robust foundation for the trial, including refining identifiers in the selection rubric sheets and

surveys to improve data matching capacity, and introducing regular data quality assurance checks.

DiD data collection

In parallel with RCT mobilisation efforts, DFNPS should start collecting data and information to deliver the DiD element of the evaluation. Specifically, historical records of colleges participating in DFNPS would need to be sourced, including which colleges participated in DFNPS, including their start and end dates of their involvement in the programme from the beginning of its delivery.

1. Introduction

The Policy Institute at King's College London (King's) was commissioned by the Youth Futures Foundation (Youth Futures) to conduct a feasibility study exploring the feasibility of a full-scale, robust impact evaluation of DFN Project SEARCH (DFNPS), a supported internship programme for young people with a learning disability, with autism, or both, who have an Education, Health, and Care plan (EHCP).

This introduction presents current evidence on the role of supported internships in improving employment for young people with a learning disability, with autism, or both, discusses the current UK approach to supporting young people in colleges with Special Educational Needs and Disabilities (SEND), such as a learning disability, autism, or both, provides a description of the DFNPS programme, and finally introduces the aims and objectives of the feasibility study.

1.1. Background

Young people aged 16-24 in the UK are facing increasingly challenging labour market conditions. Since 2002, youth unemployment has risen (Powell, 2025) and young people face longer and more complex transitions from education into work (Barnham et al., 2009). Employment plays a critical role in shaping life chances, with research showing links between individuals' physical and mental health (Peckham et al., 2019), wellbeing (Winfield and Tiggerman, 1990), and overall life satisfaction (Gedikili et al., 2021).

These challenges are particularly pronounced for young people with SEND. Evidence shows that transitions from education to the labour market result in poorer outcomes for people with autism (Ashworth et al., 2023; Jones et al., 2022) and LD (Hanson, Robinson, and Codina, 2021). In fact, these challenges persist into adulthood, as only 5.9% of adults with LD who receive support from social services are in employment (Inside Government, 2025).

1.1.1. Young people with Special Educational Needs and Disabilities

Data from the Department for Education (DfE) shows that, in 2024, around 1.3 million students in schools in England were identified as having SEND, representing 14% of the whole student population (GOV.UK, 2025). Under the current SEND legal framework in England, children and young people up to the age of 25 may receive an EHCP where their needs cannot be met through standard SEND support (Long and Roberts, 2025). In

2024, around 483,000 pupils in England held an EHCP, equivalent to 5% of the overall pupil population and approximately 27% of the pupil population with SEND (GOV.UK, 2025).

EHCPs specify the education, health and social care provision reasonably required to meet an individual's special educational needs (DfE, 2015). Local Authorities (LAs) have a statutory duty to secure the provisions identified in the EHCPs (DfE, 2015). However, while LAs are required to issue an EHCP within a statutory timeframe of 20 weeks, in practice waiting times frequently exceed this limit. In 2024, less than half (46.4%) of new EHCPs were issued within the statutory timeframe, representing a decline compared to 2023 (DfE, 2025a). These delays are attributed to different challenges such as the large amount of information required, and the limited supply of experts including educational psychologists within LAs (IES & Youth Futures, 2025).

Funding for SEND support and EHCP provision combines core education funding (such as mainstream school budgets or college place budgets) with additional top-up funding where support costs exceed specified thresholds. For post-16 learners, providers can claim Element 2 (additional SEND support) and Element 3 (high-needs top-up funding) (GOV.UK, 2025). These two funding streams are part of the wider high-needs funding system and may be used to support learners with EHCPs, although they are not equivalent to EHCP provision itself.

SEND support and EHCP funding plays a critical role in supporting young people with disabilities as they complete their education and transition into employment. However, the system is under increasing financial pressure due to rising demands for EHCPs. Total SEND spending is projected to reach £21 billion by 2029, while funding thresholds have not kept pace with inflation (Sibieta & Snape, 2024). Policy reviews highlight that while the system is widely regarded as financially unsustainable, cutting statutory entitlements risks further eroding accountability and parental trust (ibid.).

1.1.2. Transitioning into work: challenges for young people with SEND

Within an increasingly stretched support system, young people with SEND continue to face substantial challenges when they transition from education to work. Firstly, this transition involves significant environmental, psychological, and physiological changes that can increase stress and contribute to emotional and social anxiety (Malkani, 2021).

Secondly, research shows that young people with SEND face additional barriers in the labour market, including negative stereotyping, limited awareness among employers of the potential contributions of SEND employees, and insufficient understanding of the workplace adjustments and support they may require (Jacob et al., 2023). Moreover,

many young people with SEND lack access to sustained, practical support to help them secure employment and remain in work once they find a job (Hanson, Robinson and Codina, 2021).

These challenges translate into persistently poorer employment outcomes for young people with SEND compared to the general population (OECD, 2022). Young people with LD are significantly more likely to be unemployed and underemployed, and to receive low pay compared to those without SEND (Stacey & Dewey, 2024). For example, in the UK, 31% of autistic people were employed in 2022/2023, compared to 54% of all disabled people and 82% of non-disabled people, despite 77% of unemployed individuals with Autism Spectrum Disorder reporting they want to work (Abreu, 2024). Similarly, in 2023 the employment rate for people with disabilities in the UK stood at 54% compared to 82% for the general population, while less than 20% of individuals with moderate or severe LD were in employment (Powell, 2024).

1.1.3. Supported internships

Supported Internships are work-based learning placements within mainstream employment settings. They are designed to provide young people – including those with SEND – with training and work-based experience to support their transition from education into employment (Stacey & Dewey, 2024). Provision is delivered by a range of organisations, including FE colleges and third-sector providers such as DFNPS. They work in partnership with employers in the public and private sectors.

In England, supported internships are for young people aged 16–24 with SEND who have an EHCP. DfE allocates high-needs funding to local authorities, which have a duty to commission provision for these young people, including supported internship placements. Additional practical costs of in-work support, such as job coaches, are covered by the Access to Work programme (DWP). At the time of writing, there is no non-EHCP route into a DfE-funded supported internship, although reforms to the EHCP system have been signalled.

With increasing pressure on funding for this type of support, it is crucial to establish whether supported internships are an effective intervention for improving employment outcomes for young people with SEND, including a learning disability and autism, in England. Existing evidence suggests that supported internships have positive outcomes for participants, such as increased likelihood to transition into paid employment (Migliore et al., 2012), improvements in self-esteem and development of social relationships (Folguera, 2025).

A recent longitudinal study commissioned by the DfE, drawing on three cohorts of supported internship participants in England, found that almost one third (31%) of

participants had secured a job immediately after their internship, and another third (32%) were waiting to find out about a job. Six months after the end of the supported internship, almost half of participants (47%) had a job (DfE, 2024).¹

Research has also looked at specific programmes delivering supported internships. Evaluations of the DFNPS model provide strong evidence of positive effects across a range of outcomes including employment rates, health, social and emotional development, and social return on investment. For instance, a longitudinal study examining the long-term outcomes of three Project SEARCH sites in New York in the US over a five-year period found that the average employment rate was 83%, and the retention rate after one year was 82%. This was substantially higher than the average rate for other transition-to-work programmes in the region, which were 60% after one year (Christensen, 2015). This study was based on monitoring data from 124 programme participants.

Other studies found participation in supported internships yielded positive effects on participants' health, social and emotional skills (Wehman et al., 2016). An RCT of a modified Project SEARCH model – Project SEARCH plus Autism Spectrum Disorder Supports² – compared the programme with education-as-usual special education services in the US. Data were collected at four different times throughout the study, including three and 12-month follow-ups, strengthening the robustness of the findings. Among 49 high-school-aged participants with Autism Spectrum Disorder, 90% of those assigned to the supported internship group secured competitive, part-time employment three months after completing their internship, compared to 6% in the control group (Wehman et al., 2016). At 12 months post-internship, 87% of participants in the treatment group remained in employment, compared to 12% in the control group.

However, much of the existing literature have notable limitations. Many studies rely on small samples, limiting external validity, or use qualitative or cross-sectional quantitative approaches that cannot robustly estimate causal effects or assess medium- and long-term outcomes, particularly in relation to meaningful permanent employment. Additionally, the highest quality, RCT-based evidence, comes from the US, which shows the need to expand the evidence base on the effect of these supported internships in other contexts, such as England.

¹ Alongside this research, the DfE is also conducting a pilot of supported internships for young people without an EHC plan but with SEND or Learning Difficulties and/or Disabilities (LDD). At the time of writing this report, there were no publications on the findings or progress of the DfE's pilot.

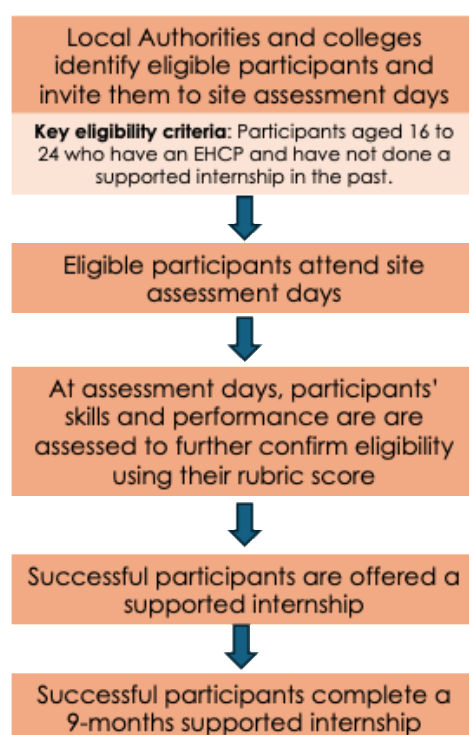
² This corresponds to the model delivered by DFNPS in England.

1.2. The DFNPS programme and selection

The focus of this study is the DFN Project SEARCH programme delivered in the UK by the charity DFN Project SEARCH (DFNPS). DFNPS provides supported internships according to their specific model, hosted by employer partners for young people aged 16-24 with a learning disability, with autism, or both, who hold an EHCP. The programme aims to equip participants with the skills, confidence, and work experience required to transition to meaningful, permanent employment, paid at a competitive wage. The programme's Theory of Change (ToC) can be found in Section 2.

This section provides a detailed description of the DFNPS delivery model. Figure 1, below, illustrates the typical participant journey through the programme, which is described in detail in the sections that follow.

Figure 1: Participant's journey through DFNPS



1.2.1. Programme structure and set up

A typical supported internship programme – referred to as a “site” – involves collaboration between different partners:

1. DFNPS staff

2. Further Education (FE) college
3. Host business
4. LA staff
5. Tutor and job coach

The roles and responsibilities of each partner is described in detail in Table 1 below. DFNPS establishes formal contractual agreements with the colleges and host businesses forming a site, which provides them with a license to operate the DFNPS programme during a specific period of time.

As of September 2025, DFNPS operated 204 DFNPS sites across the UK and Ireland, of which 155 were located in England. Of these 204 sites, 41 were in their first year of operation.

Table 1: Key partners involved in the delivery of DFNPS supported internships

Partner	Role
DFNPS staff	DFNPS staff are responsible for the overall set-up, delivery, and coordination of the supported internship model. The team is responsible for recruiting colleges and host businesses, and they work in partnership with education providers, LAs, and supported employment partners to create career pathways for young people.
FE colleges	FE colleges promote the DFNPS programme, identify participants, and refer them to the programme. They are responsible for collecting and sharing key data about participants' characteristics with DFNPS as well as monitoring their journey during and after the supported internship. Colleges will normally have an informal internal selection process, drawing on their in-depth knowledge of their learners and their needs, to decide which young people should be referred for DFNPS.

	Young people not referred may be provided with other supported internships with other providers, or recommendations on how to prepare in view of future opportunities for DFNPS selection. Some colleges host selection days at their venues.
LAs	LAs also identify young people with EHCPs who meet the programme's criteria and refer them to the programme. LAs are involved in raising awareness of supported internships (e.g., at promotional events), have responsibility for planning and commissioning funding for high-needs students, as well as issuing EHCPs.
Host businesses	Host businesses are the employers, including public (e.g., hospitals and NHS trusts) and private sector organisations (e.g., Amazon and Asda), that provide supported internship opportunities. They are responsible for ensuring that interns rotate across departments within the organisation, and that they have access to structured, ongoing support within the business.
Tutors and job coaches	Each site has one tutor and up to two job coaches, depending on the number of interns. Tutors are employed by the colleges, and job coaches are employed either by the college or by a supported employment provider. The tutor's role is to deliver the employability curriculum, and they lead on communicating with the family members, and on safeguarding concerns and practices. Tutors also support the work of the job coaches. The job coaches provide job coaching during the internship and additionally lead on conducting the vocational profiles work and on allocating interns to job rotations. They also lead on supporting young

	people with their job development and job searching towards the end of the programme.
--	---

1.2.2. Participant referrals and eligibility

Potential participants are referred to the programme by their college or LA. Before deciding whether to apply for a supported internship, candidates can attend an open day. The main purpose of the open days is to build interest in the programme, allowing candidates (frequently accompanied by their parents or carers) to learn more about DFNPS. During open days, candidates receive leaflets and brochures, and they attend information sessions detailing the opportunities offered by the programme. Open days are usually hosted by DFNPS partners such as LAs or colleges.

In order to be eligible to take part in a supported internship, individuals must meet the following eligibility criteria:

- Have a learning disability, autism, or both (based on college records, formal diagnosis is not required)
- Hold an EHCP
- Be aged 16 to 24
- Have the Right to Work in the UK
- Have not completed a supported internship in the past
- Be enrolled in the college who offers the supported internship (if young people are referred by an LA, they will nominally enrol in the college prior to start of the internship).

Additionally, it is recommended that candidates have an academic attainment between Entry Level 1 to Level 3. However, this is not a strict eligibility criterion, as each candidate is assessed individually for their suitability for the programme, and those with academic attainment outside this range may still be considered.

1.2.3. Site assessment days

Candidates that meet the above eligibility criteria can take part in site assessment days, where they undertake task-based activities and interviews, and their suitability for a supported internship is assessed by a selection team of assessors. The site assessment days referred to in this study refer to the ones conducted between March and July 2025, which is a typical timeline for assessment days selecting the forthcoming academic year cohort.

The selection team of assessors present at site assessment days is formed of representatives from key delivery partners, including the host business(es), LA(s), college(s), and job coach(es) (See Table 1). Assessors seek to establish whether the young person has the necessary communication, interpersonal, and professional skills (amongst other skills) to undertake a supported internship and adapt to a workplace environment. Assessors seek to identify those who have the necessary skills and potential to succeed in a supported internship, rather than select the highest performers, as they are likely to already have the necessary skills to secure employment independently. To guide their decision, site assessors use a range of methods and tools, including a rubric score sheet, as well as internal group discussions with other site assessors. The rubric score sheet enables site assessors to evaluate the young people on 20 different items ranging from computer skills to problem solving abilities. The rubric sheet can be found in Appendix A.

Site assessment days are typically co-hosted by a college and a host business, either at the college or a public building owned by the LA or Council. Some colleges, however, partner with more than one host business to offer supported internships in different sectors. In these cases, site assessments can take place on different days (different site assessment days for each host business), or jointly (same site assessment day for more than one host business).

Site assessment days can run in a single session or in two sessions (one in the morning and one in the afternoon), depending on the number of candidates attending. Parents or carers that accompany students either wait with students in the assessment location, wait outside in a designated space, or leave. Additionally, at some sites, parents can speak with a representative from the Department for Work and Pensions (DWP) to ask any questions they might have on how supported internships can affect their child's benefits.

Site assessment day tasks

Candidates rotate through approximately six tasks set by site assessors using a carousel system. Once candidates complete a task, they move on to the next, leaving the site assessment location once all tasks are completed.

Some common tasks include computer data entry, message relaying, interpreting maps or instructions, taking orders in a mock café setting, or performing numerical operations using calculators. However, the specific tasks vary based on the role the candidate would undertake if selected for an internship, to allow candidates to demonstrate the skills required for the roles available. For instance, candidates applying to work in a hospital may be asked to clean a wheelchair and organise hospital paperwork, whilst

others applying to work in a warehouse may be asked to pack different objects following specific instructions.

All candidates are also interviewed by a site assessor to collect general information on the candidate's life (e.g., daily routines, hygiene, independent travel skills, hobbies, and ambitions), and to help establish their suitability for the programme.

1.2.4. DFNPS supported internships

Successful candidates are offered a supported internship and if they accept the offer, they are subsequently placed with a host business. Supported internships last nine months and typically start in September. Ahead of their start date, some sites offer 'keep warm' activities through LAs and colleges, such as business visits, preparing a timetable, or training for participants to travel independently. Depending on the nature of the programme, some participants will also be subject to legal or background checks (e.g., DBS checks are required for supported internships in hospitals). Supported internships can take place in a range of different organisations, including NHS Trusts, supermarkets, retailers, and other private businesses.

Participants are unpaid during the internship, as it is characterised as part of an education or training programme which is formally recognised by the DfE. Internships provide intensive, supported learning and skill development, supported by job coaches and tutors. A typical day from 9.00am to 3.30pm would involve 1 hour in the morning of employability skills training, before then heading into the internship rotation for five hours, before concluding with reflections on the day. Supported interns are expected to rotate across three departments to gain exposure to different tasks and responsibilities. Each rotation is expected to last approximately 10 weeks, with interns receiving at least 700 hours of practical experience across the year. Participants are matched to their rotations based on their interests, skillsets, and preferred working environment. They also have an interview with their prospective department as part of preparing them to begin their rotation and to gain valuable interview experience.

While at work, participants are given support through on-the-job coaching and skills improvement strategies with the goal of achieving paid employment post-programme. Participants are supported by a mentor from the host business, but also from their job coaches and/or tutors. Mentors offer support and guidance on work-related activities, and help participants complete Vocational Profiling, a tool which allows participants to explore their ideas about workplace culture, working environment, goals, aspirations, and perceived skillset.

At the end of the internship, DFNPS encourages host businesses to organise activities to support participants, and their parents/carers prepare for the transition into employment

or another chosen destination. These activities may include preparing a summary of progress and next steps, or sharing resources to support participants with job searching.

In some instances, participants also receive follow-on support from job coaches and/or tutors up to nine months after the end of the internship. However, this type of support is not prescribed and can vary depending on the availability of job coaches and/or tutors, as well on participants' needs.

1.3. The feasibility study

Youth Futures commissioned the Policy Institute at King's to conduct a feasibility study to understand whether and how a full-scale or pilot, robust impact evaluation of DFNPS could be conducted. A secondary aim of the research was to undertake the data collection required to inform the design and feasibility of a pilot impact evaluation. This is in line with Youth Futures aims to bring about systems change for marginalised young people, to identify and generate high-quality evidence to better understand England's youth unemployment and inactivity challenge, and to learn what works in addressing this.

The feasibility study, which ran from November 2024 to October 2025, included data collection activities involving candidates who took part in site assessment days from March to July 2025, and subsequently started supported internships in September 2025.

1.3.1. Research questions

The feasibility study aimed to comprehensively explore possible impact evaluation designs to robustly assess the causal impact of DFNPS. Specifically, it aimed to determine whether a Randomised Controlled Trial (RCT) or a Quasi-Experimental Design (QED) were suitable impact methodologies. Overall, the study aimed to answer the following research questions, which evolved over time through discussion with DFNPS and Youth Futures:

Intervention model and implementation

1. What has been learned about the DFNPS model and its Theory of Change?
2. Can the intervention be improved?
3. How should business-as-usual (BAU) provision for this cohort be conceptualised?

Data collection, data quality, and infrastructure

4. Are site selection days ready to meet the data collection requirements of an evaluation?
 - What data can be collected?
 - How are data collected and stored?
 - What is the degree of consistency in the practices of data collection between sites selection days?
 - Is it feasible to collect identifiers for administrative datasets?
5. What are the most promising data sources for a future evaluation?
6. What are the characteristics of the data collected and shared by DFNPS partner sites, and what is their suitability for (i) a Regression Discontinuity Design (RDD) and (ii) a matched comparator study?
7. What data sharing infrastructure is required to ensure the availability of all required data sources for each of the specified evaluation designs?

Evaluation design and methodological feasibility

8. How is the rubric score sheet being used in practice and is it suitable as a means of identifying a comparison group for an impact evaluation?
9. What is the most promising impact evaluation design to robustly measure the impact of DFNPS?
10. Can a feasible randomised design, which is acceptable to DFNPS, be identified? If so, what resources would be required to take the design to the required scale ahead of the 2026 selection events?
11. To what extent is there likely to be scope to scale to non-participating sites, either as part of an RCT or for collecting rubric scores as part of a QED?
12. What are the key risks to the proposed design and how should these be mitigated?
13. What sample size would be required for the selected evaluation design to be able to detect the effect of DFNPS?
14. What other questions should be addressed in a future evaluation?

1.3.2. Report structure

Rather than following a question-by-question format, the report is organised around two themes: (1) programme recruitment and set up, including how DFNPS is delivered and the suitability of the programme for an impact evaluation agnostic of the specific impact evaluation design (see Section 3), and (2) the feasibility of specific impact evaluation designs that could be used to evaluate DFNPS (see Section 4). To support navigation,

Table 2 links each research question with the two themes and the concluding section of the report.

Table 2: Research questions

Research Question	Section(s)
RQ1. What has been learned about the DFNPS model and ToC?	Programme recruitment and set up (Section 3)
RQ2. Can the intervention be improved?	Programme recruitment and set up (Section 3)
RQ3. How should business-as-usual (BAU) provision for this cohort be conceptualised?	Feasibility (Section 4)
RQ4. Are sites ready to meet the data collection requirements of an evaluation? What data can be collected? How are data collected and stored? What is the degree of consistency in the practices of data collection between sites? Is it feasible to collect identifiers for administrative datasets?	Programme recruitment and set up (Section 3) Feasibility (Section 4)
RQ5. What are the most promising data sources for a future evaluation?	Feasibility (Section 4)
RQ6. What are the characteristics of the data collected and shared by DFNPS partner sites, and what is their suitability for (i) a RDD and (ii) a matched comparator study?	Feasibility (Section 4)
RQ7. What data sharing infrastructure is required to ensure the availability of all required data sources for each of the specified evaluation designs?	Feasibility (Section 4)
RQ8. How is the rubric score sheet being used in practice and is its suitable as a means of identifying a comparison group for an impact evaluation?	Programme recruitment and set up (Section 3)
RQ9. What is the most promising impact evaluation design to robustly measure the impact of DFNPS?	Feasibility (Section 4)
RQ10. Can a feasible randomised design which is acceptable to DFNPS be identified? If so, what resources	Feasibility (Section 4)

would be required to take the design to the required scale ahead of the 2026 selection events?	
RQ11. To what extent is there likely to be scope to scale to non-participating sites, either as part of an RCT or for collecting rubric scores as part of a QED?	Feasibility (Section 4)
RQ12. What are the key risks to the proposed design and how should these be mitigated?	Feasibility (Section 4)
RQ13. What sample size would be required for the selected evaluation design to be able to detect the effect of DFNPS?	Feasibility (Section 4)
RQ14. What other questions should be addressed in a future evaluation?	Feasibility (Section 4)

1.3.3. Timeline

Table 3, below, provides the timeline of the key activities and outputs of this feasibility study.

Table 3: Feasibility study timeline

Activity	Timeline
Project induction meeting completed	November 2024
Theory of Change workshop completed	December 2024
Data mapping workshop completed	January 2025
Desktop research conducted	December 2024 to January 2025
Feasibility Study Plan produced	March 2025
Site assessment days run by DFNPS	March to July 2025

Baseline surveys conducted	March to July 2025
Observations conducted	March to June 2025
Interviews with site assessors completed	April to July 2025
Impact workshop completed	July 2025
Recommendations paper produced	August 2025
Final report v1 produced	December 2025
Review and revision	January to March 2026
Final report agreed	April 2026

2. Methods

The feasibility study used a mixed-methods approach. This section discusses the quantitative and qualitative methods used in the study, the workshops organised with DFNPS staff to complement qualitative and quantitative insights, and the programme's Theory of Change.

Table 4 summarises the key data sources and collection approaches used in the study.

Table 4: data sources and collection approaches

Data source	Collection approach
Baseline survey	Surveys collected personal information from candidates and secondary outcomes at baseline before participants start their supported internship. The survey was administered by DFNPS staff at site assessment days. The survey script can be found in Appendix B.
Rubric sheet scores	Potential participants' scores were collected using the rubric score sheet, which is completed by site assessors during site assessment days. At site assessment days, site assessors collectively agree on the rubric scores after candidates have completed all tasks. The rubric score sheet can be found in Appendix A.
Monitoring data	Monitoring data from DFNPS was collected for site assessment days to measure attendance, survey completion rates, and number of candidates offered an internship.
Observations	Site assessment days were observed to better understand how candidates are assessed and selected for supported internships. Observations were conducted according to an observation guide that comprised prompts to record information around elements such as the behaviours of candidates and staff members, the characteristics of the environment, and the way candidates engaged with each other and with evaluators.
Site assessor interviews	Interviews with site assessors were held online and covered questions on how candidates are assessed, how they collect data on

	candidates, and broader questions on the perceived feasibility of an impact evaluation.
--	---

These data sources were complemented with three key workshops:

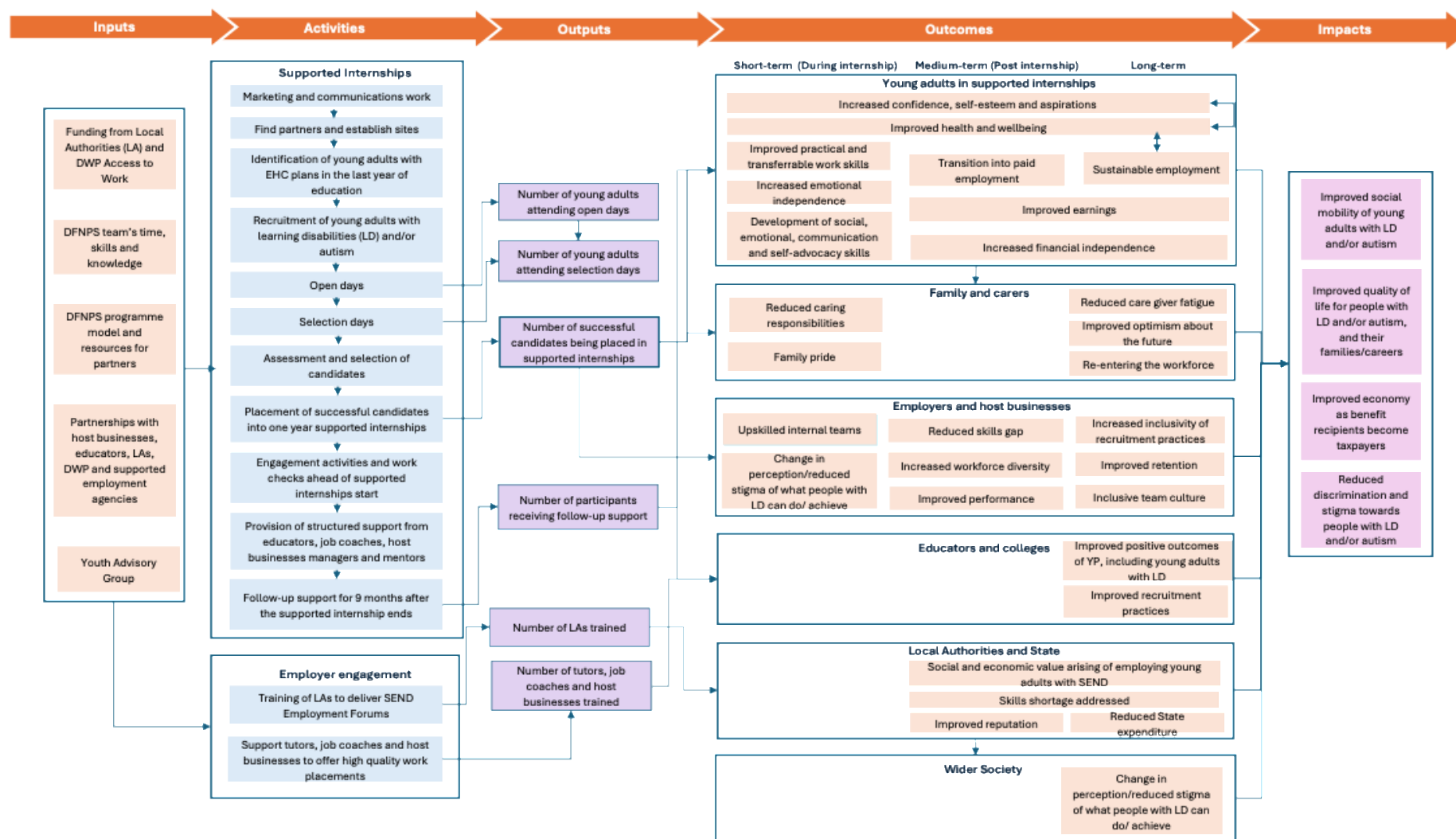
- Theory of Change workshop:** In December 2024, an online workshop with key DFNPS staff was held to design the programme's Theory of Change and gather key insights on how the programme works and what are its anticipated outcomes.
- Data mapping workshop:** Various members of DFNPS attended an online data mapping workshop in January 2025. The aim of the workshop was to review and identify the key primary and secondary outcomes of interest, review existing data collection processes and capacity, and explore what data could be collected during site assessment days. The insights of this workshop were used to develop a data collection framework.
- Impact workshop:** In July 2025, a workshop with key senior members of DFNPS was held to explore the feasibility and acceptability of different impact evaluation methodologies. This workshop was used to draft a Recommendations Paper for Youth Futures, with early insights on the feasibility of different RCT and QED options, which then fed into the prioritisation of remaining activities for the feasibility study. The workshop was followed by a call with the programme's CEO to capture their views and answer any potential questions around each methodology.

2.1. Theory of Change

The ToC diagram presented in Figure 2, below, shows the programme's primary beneficiaries and anticipated impact over time. Appendix C also outlines the causal mechanisms through which the programme is expected to achieve its intended outcomes, as well as the key risks and assumptions that underpin the programme's ToC.

As shown in the diagram, DFNPS aims to support young people with a learning disability, with autism, or both, to achieve full-time jobs that lead to rewarding careers in integrated settings, paid at a competitive wage. Through meaningful exposure to employment, DFNPS ultimately aims to improve the employment and wider outcomes of young people with a learning disability, with autism, or both, as well as improve their own and their families' / carers' life quality. More widely, DFNPS supported internships aims to contribute to an improved economy as benefit recipients become taxpayers (or they receive less benefits related to employment status), and to reduced discrimination and stigma towards people with a learning disability, with autism, or both.

Figure 2: DFN Project SEARCH Theory of Change



2.2. Data collection and analysis

2.2.1. Quantitative data

This report uses quantitative information gathered through three data sources: monitoring data capturing site attendance levels and survey completion rates, a survey of DFNPS candidates, and the records on candidates' scores from the rubric sheets. While it was not possible to fully merge participants across all these data sources, we provide a discussion of the specificities of each dataset and the merged sample in the following sections.

Monitoring data

King's received data compiled by DFNPS to capture the characteristics of site assessment days. These included the number of attendees, the number of candidates deemed eligible for the supported internship (based on their performance at site assessment days), and the number of candidates who refused to complete the survey (see section below on candidates' survey). Data were shared with King's through an online tracker, which DFNPS staff regularly updated throughout the duration of the programme.

In total, the DFNPS system recorded these data for 37 site assessment dates, out of 68 organised in 2025, with the dataset including information covering 349 candidates. The number of candidates recorded in this monitoring dataset is lower than the ones recorded in the candidates' survey and rubric sheets (see following sections). This reveals some of the detected issues regarding data completeness that should be prepared for in a full-scale evaluation. It is also important to highlight that these records include all attendees to site assessment days, a group that includes candidates that could have refused to take part in the study or who could be not eligible for the internships.

Candidates' survey

A survey of candidates during site assessment days was used to capture data on their personal characteristics, secondary outcomes at baseline, and general identifiers for linkage with administrative data, including name, college, date of birth, postcode, and

email address. Linkage architecture was set up so that the feasibility study's data could be used as a pilot dependent on the pertinence of specific evaluation designs³.

As part of the survey development process, measures of secondary outcomes were developed through online cognitive testing interviews with young people with SEND, to make sure the survey was easy to complete, clear, and that the questions captured the concepts of interest for the research. As part of cognitive testing, young people were asked to "think aloud" as they answered the questions to show their thought processes. Additionally, interviewer observations and verbal probing were used to explore specific aspects of the questions and test for comprehension, interpretation, and the suitability of response options. The finalised survey script can be found in Appendix B.

The survey was administered by DFNPS staff members that attended site assessment days (referred to as "associates"). Early difficulties, such as low participation rates, were quickly resolved by integrating the survey into the rotation of activities that young people did during the site assessment days. Following an early site assessment day where only 6 of the 18 young people in attendance participated in the survey, DFNPS associates were able to survey all attending young people that consented to it at the following site assessment days, except for two candidates who were not able to complete it. It is worth noting that there was large variability around when site assessment days were scheduled, with some of the most attended taking place early on in April and May 2025.

DFNPS associates were trained by the King's evaluation team on how to obtain informed consent and how to conduct surveys with candidates. DFNPS staff briefed candidates on the survey's aims using accessible information sheets, and only those who provided informed consent proceeded to complete the survey.

In total, 406 candidates completed the survey. Indicative consent-to-participate and response rates have been estimated based on the available record from the monitoring data provided by DFNPS, which is discussed above. Based on these data, 81% of candidates consented to complete the survey, and all those who consented completed it. Of the remaining candidates, 12 did not consent to participate and a further 54 did not complete the survey where details were not recorded as to why, but we hypothesise these were also declines to participate in the research that were not consistently recorded in the monitoring data.

³ If conducting an RDD were feasible, data from DFNPS's 2025 cohort would have been used as first year of analysis, to be top up with subsequent cohorts to reach suitable sample sizes.

Rubric sheets data

To capture candidates' suitability for the programme, DFNPS associates collected the rubric scores of 399 candidates and shared it securely with the King's evaluation team. The difference between these figures and the candidates' survey records may be due to gaps in information collected from colleges. The rubric sheets also collected additional information on whether potential participants were offered an internship after the assessment. This question, however, was added after data collection had already started, with data only available for 179 out of 399 candidates for which DFNPS collected rubric sheet scores.

Rubric sheets only contained the name and site as identifiers for the candidates, which complicated the linking of this data with other data sources. It is recommended that additional identifiers are included in a future full-scale evaluation.

Merged data

All three quantitative data sources (monitoring data, baseline survey, and rubric sheet scores) were harmonised to facilitate the analysis. Table 5 provides an overview of the results of the merging. First, data from candidates' surveys (406 unique responses) were merged with data from rubric sheets (399 unique responses) using general identifiers including name and site. Inconsistencies were also manually clarified with DFNPS' support to maximise merging rates.

In total, 377 candidates (93% of all survey responses) had both a survey response and a score record from the rubric sheets, leaving 24 unmatched rubric scores and 37 unmatched survey responses. The monitoring data remains a third source of data overlapping both with survey data and rubric sheet data, but as records were only available for 37 of the 68 site assessment days which ran in 2025, these records had high levels of incompleteness. Monitoring data were still used to estimate an indicative survey response rate, and investigate variation in site assessment days characteristics.

Table 5: Data sources

Data Sources	Number of unique responses
Candidates' survey	406
Candidates' rubric scores	399
Merged candidates' survey and rubric	377

Data Sources	Number of unique responses
Unmatched survey responses	37
Unmatched rubric scores	24
Monitoring data (site assessment days)	37 out of 68 organised
Attendance to site assessment days*	349
Surveys completed at site assessment days*	283

* Incomplete data due to data from only 37/68 of site assessment days recorded, leading to the data not matching survey and rubric data.

Quantitative data analysis

Quantitative analysis was conducted using the merged survey and rubric dataset in R to examine patterns of candidate characteristics, engagement with site assessment days, candidates' rubric scores and internship offers. Analysis included both descriptive and inferential investigation of the data, with subgroup analysis to highlight differential patterns. Descriptive analysis included generating descriptive statistics of central tendency and dispersion, visualising the key variables that describe the sample demographically and with regards to programme participation. Inferential analysis used regression models to identify significant variation by participants' characteristics in key variables of interest, including their rubric scores.

The analysis is presented as a way to understand in more depth how DFNPS selection ran in 2025, with the aim of generating new evidence to inform the feasibility of conducting a future impact evaluation, including by comparing the rubric scores and demographic characteristics of candidates offered for a supported internship with those not. The results of this analysis are presented next in Section 2.2.2.

2.2.2. Qualitative data

The qualitative data collection consisted of observations of site assessment days and interviews with site assessors. For both, researchers worked closely with DFNPS to identify a suitable set of selection days to observe, which allowed them to capture the variation in local contexts and employers, as well as identifying a sample of interviewees.

Observations

Researchers conducted seven observations of site assessment days out of the 68 total organised (~10%) in 2025 to better understand how candidates were assessed and selected for DFNPS. Observations also allowed researchers to gather insights on how the rubric was used by site assessors, how sites were organised, and how the survey was administered in practice.

Table 6 below summarises the sites that were observed as part of this study. Researchers attended sites across different regions in England between March and June 2025. Each observed site involved distinct host businesses and varied in the number of candidates participating in the assessment days (between a minimum of 10 and a maximum of 30). Of the seven sites, five held joint assessment days (i.e. candidates were considered for multiple host businesses at once), with representatives from each organisation present. Observations were conducted according to an observation guide that comprises different prompts in relation to both candidates and staff members, such as the environment, body language and tone, candidates' engagement, materials used, and agency. The observation guide can be found in Appendix D.

Table 6: Sites observed

Employer for the site	Date	Location	Number of candidates
Great Western Hospitals NHS Foundation Trust; Amazon; Swindon Council	04/03/25	Swindon	18
Lambeth Council; NHS England; Wellington House; Kings College Hospital NHS Foundation Trust	25/03/25	London	25
Greenwich Council; Berkeley Homes; Queen Elizabeth Hospital	24/04/25	London	13
ISS (at Union Bank of Switzerland); Queen Mary University London	09/05/25	London	12
NHS Blood and Transplant	21/05/25	Bristol	16
Royal Lancaster Infirmary	22/05/25	Lancaster	10
Chartwells at De Montfort University; Leicester Marriot Hotel; University Hospitals Leicester Trust – Glenfield Hospital.	13/06/25	Leicester	30

Interviews with site assessors

Researchers conducted 19 semi-structured interviews with site assessors. All interviews were completed online and lasted up to 60 minutes. They covered questions on how selection days are organised, how candidates are assessed, how they collect data on candidates, as well as broader questions related to the perceived feasibility of an impact evaluation.

As shown in Table 7, interviewees were selected among staff from colleges, host businesses, and LAs who were involved in site assessment days. Interviewees were recruited during observations, with King's researchers collecting contact details of site assessors that were happy to be interviewed afterwards. To broaden the sample of the interviews and capture a wider variety of experiences and views, further assessors were recruited by the DFNPS team, who extended the invite to a range of site assessors from different sites that were not observed.

Through observations, researchers found that there were approximately five site assessors per site assessment day, and some site assessors might attend more than one site assessment day. With 68 site assessment days conducted, this implies having interviewed at least approximately 6% of site assessors.

Table 7: Partners interviewed

Partner type	Number of interviewees
College staff	9
Host business staff	5
LA staff	3
Job coach	2

Qualitative data analysis

Interviews were recorded and transcribed by a professional transcription company. All transcripts and observation notes were entered in NVivo 14 for data management and analysis. Data from the interviews and observations were managed using a thematic framework matrix. This approach allowed for the identification of commonalities and differences in the diverse qualitative data collected and to create descriptive

categories and explanatory concepts (Gale et al., 2013). Researchers then conducted explanatory analysis to look for linkages, associations and possible explanatory concepts that will provide a rich understanding of the programme (Ritchie et al., 2014).

2.3. Feasibility assessment

This feasibility report considers in-depth four overarching impact evaluation designs to deliver a high-quality, robust causal impact evaluation of DFNPS: two RCT designs, a Difference in Differences design (DiD), and a Regression Discontinuity Design (RDD).

Five progression criteria were identified and agreed with Youth Futures to determine feasibility for full evaluation. The criteria, initially agreed as part of the study's feasibility plan, were further refined to ensure it accurately captured the different elements that determine the feasibility of each impact methodology. The final progression criteria are:

1. **Robustness:** the strength of the design in generating high-quality evidence;
2. **Practical feasibility:** the practical viability of implementing the impact design, including related to time and resource constraints;
3. **Acceptability:** the extent to which DFNPS, King's, and colleges support the design;
4. **Working relationships:** the existence and quality of necessary relationships between King's, Youth Futures, DFNPS, colleges, and other key stakeholders including DfE, LAs, and host businesses;
5. **Ethical considerations:** ethical considerations and feasibility of addressing or mitigating these.

The progression criteria, further detailed in Appendix E, are assessed based on the qualitative and quantitative data sources analysed. Additionally, during the impact workshop, acceptability of each design was also discussed with DFNPS staff. Furthermore, assessment was also informed by conversations with college representatives, particularly those involved in the recruitment of participants for supported internships schemes and data management, to better understand the operational feasibility of the different designs.

A key element of the practical feasibility criterion, which assesses the practical viability of implementing each impact design, also relates to the feasibility of evaluation timelines. Youth Futures is only able to fund evaluations that will conclude within its current endowment period, which ends at the end of 2028. As such, this consideration is also

explored to assess the feasibility of Youth Futures funding each impact method following this feasibility study within this report.

For each potential design, each progression criterion was given a RAG (red, amber, or green) rating – **green** if the criteria show full readiness for a full-scale evaluation or a clear path to full readiness, **amber** if reasonable changes are required but achievable, and **red** if the changes required are either considerable or not at all achievable in an appropriate timeline. Further details on the RAG rating approach can be found in Appendix E.

In August 2025, we produced an internal Recommendations Paper for Youth Futures, which outlined our early reflections on the feasibility of conducting a full-scale impact evaluation of DFNPS, and the specific designs under consideration. These reflections were based on the feasibility study activities undertaken to date, which included several of the qualitative and quantitative research activities as described. The Recommendations Paper was already able to conclude that several impact evaluation designs, which can be found in Appendix F, were infeasible. This final report builds on that work and presents a final set of evaluation designs which were assessed further in more depth, which could feasibly be taken forward to evaluate the impact of DFNPS.

2.4. Key methodological limitations

2.4.1. Triangulation of data sources and missing data

This study relies on quantitative data from multiple sources (surveys, rubric sheets, and monitoring data). Triangulated data strengthened the analytical capacity of this research, but its use also resulted in some notable issues.

First, as outlined earlier in this section, the coverage of all three data sources is incomplete. Not all survey participants have a rubric sheet record, and vice-versa, and only approximately 54% of site assessment days were included in the monitoring data, as outlined in Table 5. Based on the number of survey responses, the largest of the databases used in this report, 93% (377) of respondents could be matched to a rubric sheet record, and 70% (283) were captured in the monitoring data. This is compounded by the challenge with accessing records around internships offers and take up rates, resulting in 55% of these data entries being missing which would be the main indicator of whether a young person is effectively engaging with the supported internship or not.

Incomplete records affect the analysis of candidates' demographic and survey data, and how that relates to scores from the rubric sheet and supported internship offers. The incomplete monitoring data also affected the calculation of survey response rates.

Response rates should be considered cautiously. Missing data around offers take up is particularly concerning. For any large evaluation, it is key to access the take up rate that the programme may have, as this affects the necessary power to detect an effect, highlights the need of including compliance analysis and allows further exploration of attrition patterns, which constitutes a limitation.

A further complication is the presence of eight cases where an individual attends multiple selection days. For the purposes of inferential analysis, these cases were treated as separate individuals, but any descriptive statistics collapse the multiple records down to the single original survey, rubric score, or site assessment day record.

The support of DFNPS was crucial in efforts to increase the extent to which high quality data could be collected and data sources could be merged, vastly improving the quality of the data in the process of collecting it and analysing it. A number of adaptations were made during the feasibility phase to streamline data collection for the DFNPS team. These included changes to how recruitment for interviews took place, and how candidates' surveys were collected during site assessment days, but also to the way site assessors reported the rubric scores, and what data were collected within the rubric sheets themselves and within the monitoring data environment. Future research and evaluations should take these challenges into account and consider approaches such as unique identifiers and more robust data collection practices to limit harmonisation issues and improve the quality of the data available.

2.4.2. Rubric sheet design and rubric data collection

Scores from the rubric sheet were initially collected using Qualtrics but collection later switched to Excel spreadsheets due to the difficulty of accessing multiple candidates' surveys on Qualtrics. The use of Excel introduced inconsistencies in reporting, as assessors did not adopt consistent response approaches (both across assessors within site assessment days, and across site assessment days), which complicated the matching of records and required additional verification. A further limitation of the rubric sheet is that it did not contain any identifiable data apart from the candidate's name, entered by the site assessors, often with erroneous spellings, which complicated the data merging process as well.

DFNPS was once again proactive in seeking to improve the rubric sheet data records, assisting researchers by verifying matches or finding missing rubrics. Future rubric scoring spreadsheets and data collection approaches should include more identifiable data, such as the participants' college and their date of birth, which combined with name records can improve the ability of researchers to correctly match rubric sheet and survey data.

2.4.3. Interviewees self-selection bias

A notable limitation of the study concerns the potential for self-selection bias among interview participants. Because interviews were conducted with individuals who volunteered to take part and who were identified by DFNPS, the sample was not randomly selected. This means the findings may not fully reflect the views of all those involved in or affected by the DFNPS programme. While self-selection can sometimes lead to disproportionate participation from people with particularly strong views—positive or negative—we did not observe a consistent pattern in one direction during the interviews. Rather, participants offered a range of perspectives. The perspectives presented in this report should be interpreted as those of the individuals who agreed to be interviewed, rather than a definitive representation of all potential participants.

2.4.1. Lack of research on supported internships themselves

The feasibility study did not include any research on the nine-month supported internships themselves; it only focused on the selection process as this is where randomisation would take place or features of the existing process would need to be accounted for in an impact evaluation design, and there is much less existing evidence on how participants are selected compared to how supported internships are delivered. As such, the feasibility study did not explore delivery and fidelity of the supported internships or fully investigate data collection at endline; neither any recruitment and onboarding process between site assessment day and internship start.

3. Programme recruitment and set up

This section presents findings and insights on programme recruitment and set up, drawn from observations and interviews with site assessors, alongside an overview of attendance patterns, consent rates, and the profile of research participants. It especially addresses RQ 1, 2, 4 and 8, identifying key learnings on the delivery of DFNPS, and discussing the key enablers that lay the foundation for a future impact evaluation, but also some of the potential challenges that would need to be addressed to deliver a robust evaluation. The remaining RQs around the feasibility of specific evaluation designs are addressed in Section 4.

3.1. Site assessment days

Upon referral from colleges or LAs, and after attending the optional informative open days, candidates' first interaction with DFNPS is at site assessment days. Findings from observations show that site assessment days are run on a site-by-site basis, varying significantly in their format. In this section, we will explore this variation and provide further insights on the running of these days beyond the intended approach previously described.

3.1.1. Venue set up

Observations of site assessment days confirm that the format of, and space for, site assessment days varied considerably. For instance, some were organised in large venues such as LA buildings or large college rooms, where most activities were delivered in one large room. Other venues required activities to be organised in several rooms. Generally, candidates' interviews took place in a quiet, separate space.

King's interviews with site assessors identified some notable frustrations, highlighting that not all site assessment days were held in an adequate venue. For instance, some did not have enough space, and so candidates had to perform activities and interviews in front of other candidates, which inhibited them. Moreover, the lack of space sometimes meant that it was difficult to have a separate waiting area for parents, and site assessors reported that some parents wanted to accompany their children during the tasks, which could interfere with their performance.

These findings highlight the importance of securing appropriate venues to ensure space does not influence candidates' comfort and performance. Ensuring that assessment

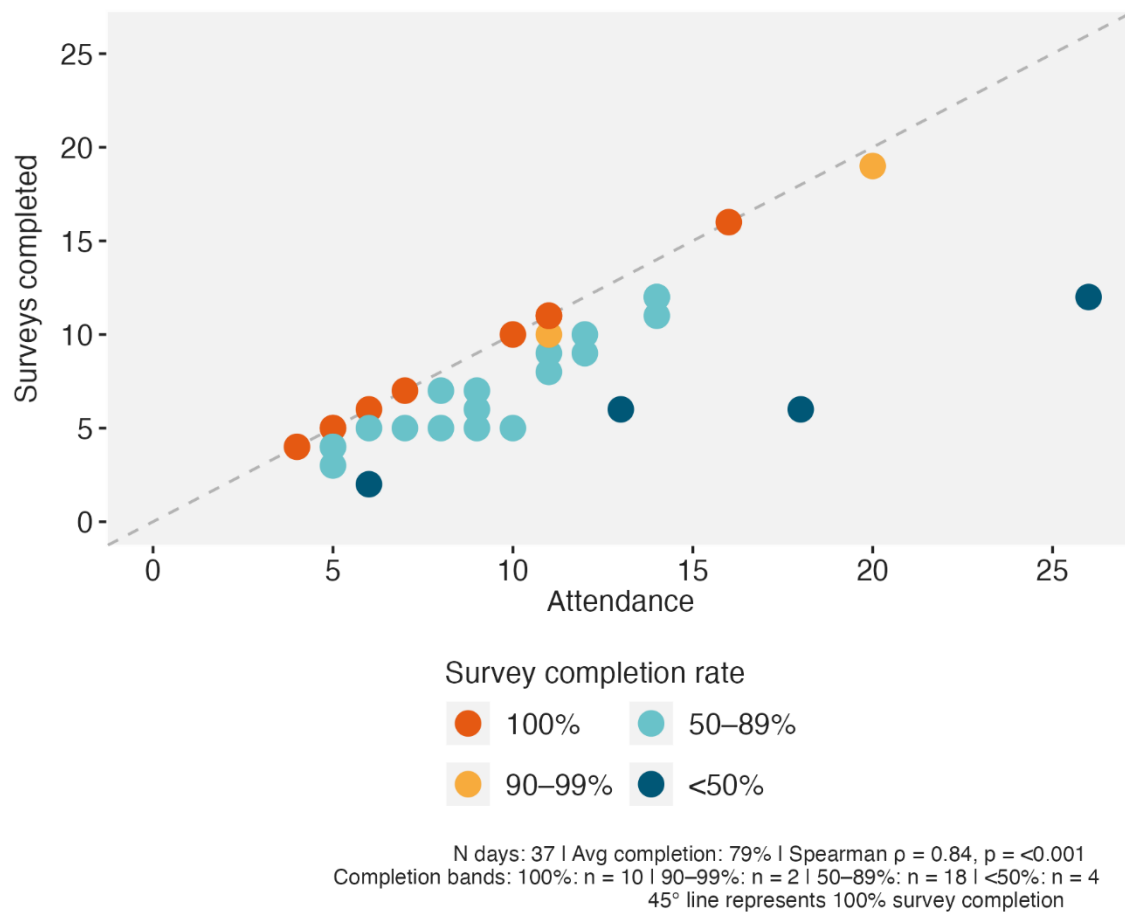
activities and interviews can take place in quiet, private areas would be essential for maintaining consistency across sites and reducing contextual bias in a future impact evaluation. Likewise, providing a clearly designated waiting area for parents or carers is crucial to prevent unintentional involvement that could affect young people's responses or the integrity of the baseline data collection process.

3.1.2. Attendance levels

Attendance levels were another factor varying substantially across site assessment days. Overall, monitoring data included details about 349 candidates who attended one of 37 site assessment days, and attendance varied between the smallest site assessment day with four candidates, and the largest one with 26 candidates.

In addition to monitoring overall attendance levels, DFNPS monitoring data also captured survey response rates, allowing us to assess the relationship between the two, illustrated in Figure 3. The 45° dashed line represents cases where 100% of candidates completed the survey. For most selection days completion rates were near 100%. Descriptively, completions rates are noticeably lower for two of the four site assessment days with the highest attendance, pointing to the importance in future studies to monitor survey responses rates particularly carefully for larger assessment days.

Figure 3: Correlation between site assessment day attendance and survey completion rate



3.1.1. Survey completion and consent rates

Overall, 81% of candidates accounted for in the monitoring data consented to and completed the survey, showing that candidates were willing to take part in the research, but also that no undue pressure was put on them to answer the survey as a significant minority felt comfortable to not consent. Administering the survey also showed a test case for effective collaboration with DFNPS to obtain consent from candidates and gather their survey responses. A key enabler of this process was DFNPS associates administering the survey as part of the assessment day activities. As explained in Section 2.2.1 on candidates' survey, DFNPS associates received training from the King's evaluation team on obtaining informed consent and conducting the survey, and candidates were briefed individually using accessible information sheets before deciding whether to participate.

3.1.2. Assessment of candidates

In general, observations showed that the site assessment day tasks were within most candidates' capability range and allowed them to confidently demonstrate their capacity, albeit with the support of the assessor on occasions. This suggests that the college's pre-screening process conducted before referring young people to the site assessment days was successful in identifying suitable candidates. From an evaluation perspective, this means though that the pre-screening also reduced the scale of a possible comparison group to be identified at site-assessment days.

The King's evaluation team observed that site assessors were responsive to candidates' needs, provided support when candidates seemed nervous or unsure on how to perform a task, and were friendly with candidates to ensure they were as comfortable as possible. Overall, the assessors were well-trained and enabled accurate assessment of suitability for the supported internships.

Upon completion of all tasks by all candidates attending a session, site assessors discuss the suitability of each candidate in turn. The discussion takes place either during site assessment days (in-between the morning and afternoon sessions, or at the end of the day), or during a later follow-up opportunity. As confirmed during observations, during these meetings, assessors considered the performance and suitability of each candidate, sharing the rubric sheet scores they assigned to each candidate based on their performance in the tasks, and agreeing on an overall score for each of the rubric sheet elements.

Use of the rubric sheet

Key to the assessment of candidates are the rubric sheet scores collected by the assessors. Overall, interviews and observations show that the rubric sheet is a useful tool to provide a degree of consistency in assessors' choices, with site assessors generally agreeing that it is comprehensive and useful. Interviewed site assessors expressed satisfaction for having a framework to refer to, and some site assessors who had to explain the selection outcome to families and candidates found it is a valuable resource, with additional evidence, to explain their decision and inform their communication approaches.

Feedback collected from site assessors suggests that during the assessment meetings, the rubric was most useful to highlight specific concerns, such as independent travel skills or self-care. This meant that individual markers were, in some instances, of more interest than overall or average rubric scores. This links to the ToC, which highlights DFNPS' efforts to match students to suitable host businesses that might require specific skills and attributes. The rubric also appeared appropriate for assessing the key tasks and offered

sufficient flexibility for site assessors. Assessors were able to leave a rubric score blank when it was not applicable to the activity they were observing. For example, if a task required a candidate to communicate a message to someone, the assessor would not complete the computer-skills section of the rubric, as they would have no basis on which to evaluate that skill. Instead, this score would be provided by a different assessor responsible for a task where computer skills were directly observed.

Interviews with site assessors and observations also highlight some challenges when it came to the use of rubric sheets. First, while most assessors showed they understood how the rubric sheet worked, some reported there had been cases when unfamiliarity with the scoring system and misunderstandings arose. For instance, one interviewee reported that the use and interpretation of the rubric sheet seemed arbitrary, recalling that in some instances candidates were marked down according to how many prompts they needed rather than in reflection of their performance. In a few instances, some site assessors felt that the rubric was being enforced to support the research and that it did not improve on the previous ways of choosing candidates based on personal testimonies and broader discussions.

Observations show that final rubric scores were also sometimes interpreted through descriptive language (e.g., "all 4s means he did really well"), but this depended on site assessors' individual approaches and were not standardised across site assessment days, reflecting a lack of standardised guidance on how to assign final rubric scores. There were also disagreements regarding the treatment of high-scoring candidates, with some assessors often questioning whether an unpaid supported internship was right for someone who potentially is sufficiently skilled to apply for other internships or find employment without DFNPS support, whereas assessors from host businesses often wanted to select the highest/ very high-scoring candidates.

Given these findings, it is important to be cautious in relying on the rubric score as a core component of any evaluation. It is used differently across different sites and is more a starting point than a single point of truth for whether a candidate is successful in obtaining an internship. It is our view that an impact evaluation should look to incorporate the rubric as it is used (i.e. as one component of understanding whether a young person is suitable) rather than looking to change processes around the rubric (e.g. standardising use, setting a cut-off) as this is likely to be difficult to ensure fidelity to and impacts external validity of the findings.

Site assessors' discussion

On top of discussions around rubric scores, site assessors also tend to incorporate wider informal insights into their assessment of the suitability of candidates, including reflections on interview responses, or reflections on candidates' behaviour in college. Taking these

other insights together with the rubric scores, each candidate is then assessed as suitable for a supported internship, not suitable, or maybe suitable with further information needed (e.g., information on attendance might not be available straightaway so a decision is put on hold).

Site assessors agreed that the collective discussions about students' suitability work very well, despite the length of the day. It is worth noting that some site assessors saw site assessment days as exceptionally long, as the set-up tends to start before 9am and the day concludes after 5pm.

Discussions seemed more effective when split, one after the morning and another after the afternoon session. There was some lost accuracy if the selection team tried to discuss all candidates at once, as some site assessors forgot certain candidates rubric scores or performance cues. Some proposed using candidates' pictures as a prompt, depending on the availability of the right support technology. Nonetheless, the host businesses interviewed appreciated that they got a say in the selection of students that they take on, improving their sense of agency and engagement in the programme.

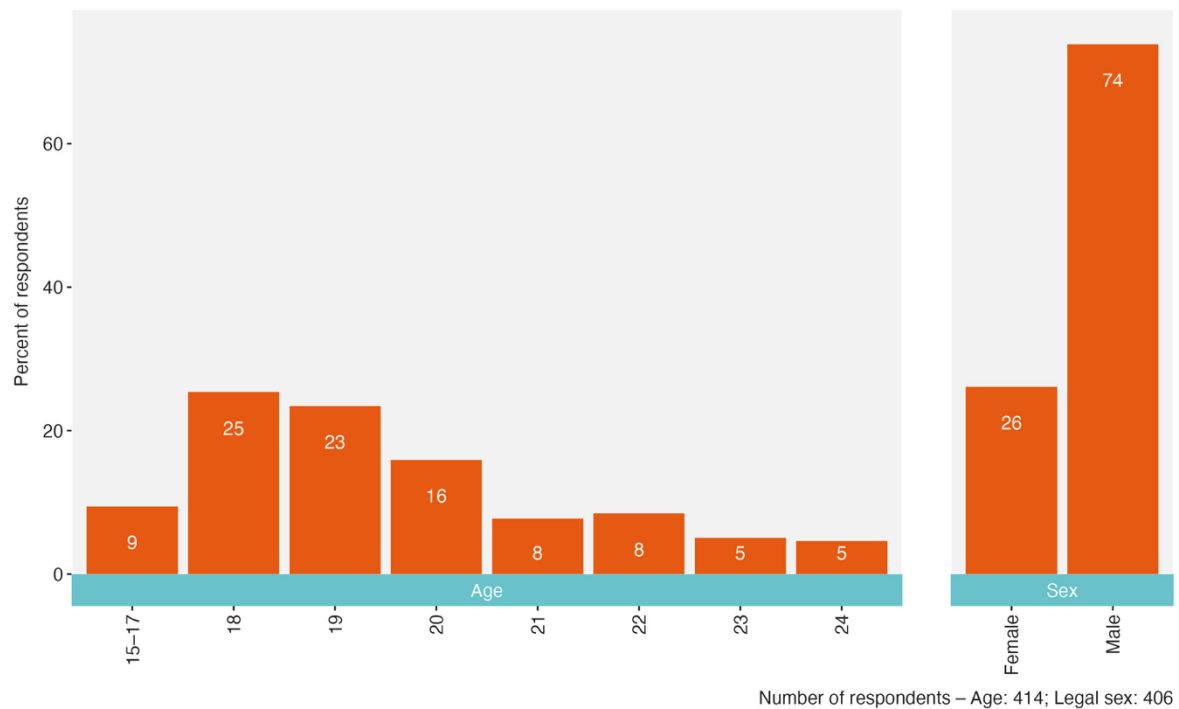
3.2. Candidates' characteristics and outcomes

This section uses data from the survey, rubric sheets, and site assessment monitoring to summarise candidates' demographic characteristics, and their outcomes from site assessment days, including their rubric scores and offers for supported internships.

3.2.1. Sociodemographic characteristics

Error! Reference source not found. shows the distribution of the key sociodemographic variables captured in the candidates' survey. Respondents had a mean age of 19.5 (with 60% of all survey respondents aged between 18 and 20). The distribution skewed towards men, with 73% of all respondents being male, with 2% (n=10) of the whole sample also indicating their gender identify differed from their sex assigned at birth.

Figure 4: Survey respondents' demographics

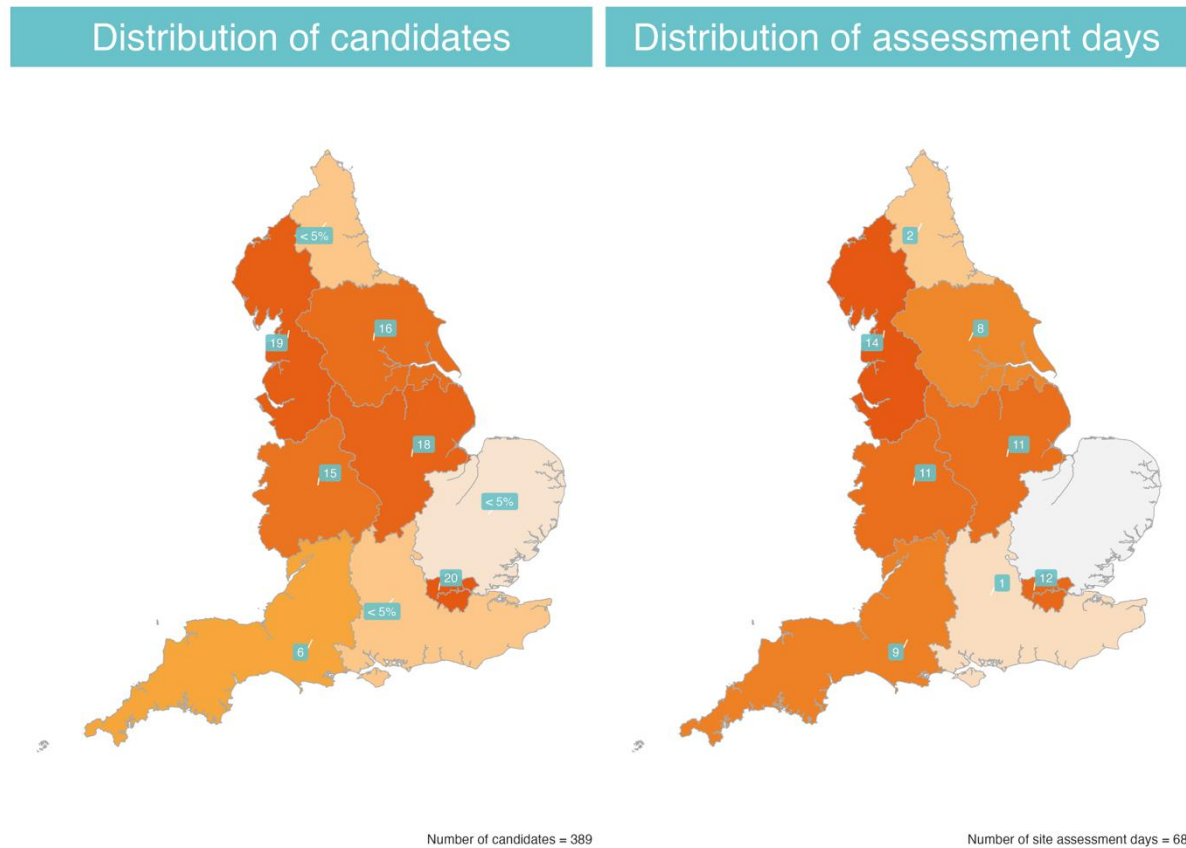


As shown in Figure 4, there were some notable disparities in the regional distribution of survey respondents across England, as presented in Figure 5, with 20% responses coming from candidates in London, followed by 19% in the North West and 18% in the East Midlands, compared to less than 5% candidates from candidates in the East of England, the South-East, and from the North East.

Finally, 25% of respondents said they had previous experience of paid work before attending the site assessment days. Of these, a follow-up open-ended question shows a mix of responses either detailing the type of employer (supermarkets, pubs, cafes, shops) or the nature of the employment (temporary work, voluntary or unpaid work, even though the question specified paid work experience, illustrating the need in a future evaluation to be specific in questions about employment).

Figure 4: Distribution of candidates and site assessment days across England regions

Figure 4: Distribution of candidates and site assessment days across England regions



3.2.2. Rubric sheet scores and supported internship offers

Table 8 describes candidates' rubric scores for the overall sample, those who received an offer for a supported internship, and those who accepted the offer.

Overall, with possible rubric scores between 20 to 100, the mean score is 78, and the median is 80. The lowest observed score is 37, a clear outlier, with the remaining 99% of the sample scoring 50 or higher. Based on the reduced sample of 179 of the 399 candidates (45%) which had data on internship offers, 154 (86%) candidates received a supported internship offer, 22 (12%) did not receive an offer, and the remaining four (2%) received conditional offers or were waiting for further information at the time the data was collected (e.g., conditional on passing a DBS or background check).

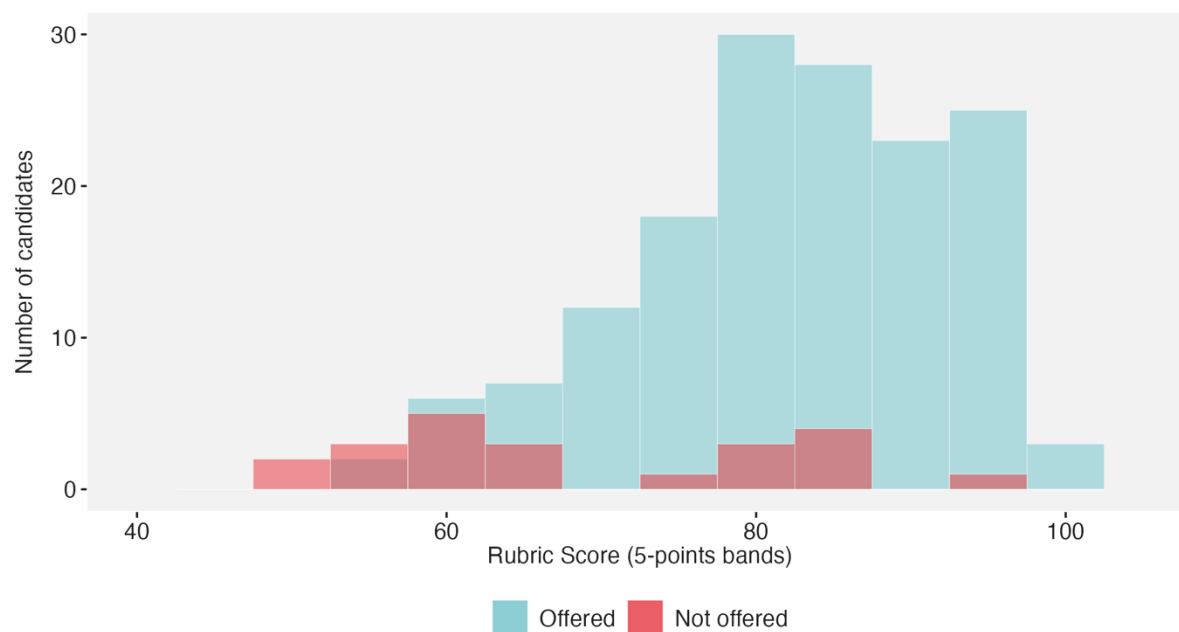
Table 8: Descriptive statistics for candidates' rubric scores from the rubric sheet

Group	Min	Max	Mean	Median	SD	n
All candidates	37 ⁴	100	78	80	12	399
Offered a supported internship	55	100	82	83	10	154
Not offered a supported internship	51	93	69	63	13	22
Accepted a supported internship	55	98	82	82	10	115

Figure 5 shows the score distributions for those offered or not offered the supported internship. Notably, both those with and without supported internship offers have a similar range of scores, but average scores are higher for those with offers (82) compared to those without (69). This difference is statistically significant and holds after controlling for respondents' demographic characteristics. In terms of predictors of the likelihood of receiving an offer, none of the observed demographic characteristics were significantly associated with it; the main predictors were the rubric score (as expected) and, to a lesser extent, the reported baseline measure of confidence. There is also no evidence of a significant difference between those who were offered and those who accepted supported internships, though this may reflect small sample sizes rather than the absence of a true underlying difference. The regressions used to explore these associations can be found in Appendix I.

⁴ Note that for this individual there was no information available on whether they received or had accepted an internship offer by the time this report was written.

Figure 5: Distribution of rubric scores for candidates offered vs not offered DFNPS



Offered n = 154 | Not offered n = 22

Offer data for candidates with rubric scores under 50 were not available

4. Feasibility assessment

This section presents the feasibility assessment of four different evaluation designs that were considered as part of this feasibility study. It especially addresses RQ 3-5 and 7-14. We first consider the outcomes and possible data sources for the evaluation, as this is a core component of feasibility across all possible designs. Each subsequent subsection then considers the evaluation designs in turn, focusing on enablers and barriers and exploring different progression criteria (robustness, practical feasibility, acceptability, working relationships, and ethics). This is followed by an overarching discussion providing our recommendations for a future impact evaluation.

Table 9 summarises the RAG assessment for the impact evaluation designs. We also considered three further designs, but these were deemed not acceptable or feasible at a relatively early stage of the study (in the Recommendations Paper), and so are covered only briefly in Appendix F.

Table 9: Summary of the feasibility of different impact evaluation methods

Approach	Description	Robustness	Practical feasibility	Acceptability	Working relationships	Ethics
Individual-level RCT with scale up to SEND participants without EHCPs	Programme delivery is expanded to offer new supported internships to SEND students without EHCPs. Participants within the trial are randomised into a treatment (DFNPS) and control (BAU) group, which are compared to identify the impact of DFNPS.					
Individual-level RCT with scale-up to new colleges	Programme delivery is expanded to offer new internships to students with EHCPs recruited in a new set of colleges. Participants within the trial are randomised into a treatment (DFNPS) and control (BAU) group, which are compared to identify the impact of DFNPS.					
Difference-in-Difference (college-level)	Historical administrative data is used to compare the aggregate employment outcomes for young people who hold an EHCP within DFNPS-participating colleges to similar young people within a matched comparator set of colleges not participating in DFNPS, considering the time-period when each college joined DFNPS. This is used to identify the impact of DFNPS.					
Regression Discontinuity Design (2025 cohort, or 2025 and 2026 cohorts)	Using 2025 cohort data and exploiting the element of rubric scores which influences selection for DFNPS, participants above and below the cut-off point that are and are not offered a supported internship are compared to identify the impact of DFNPS.					

4.1. Outcomes and data sources

Identifying appropriate outcomes and how they will be collected is fundamental to the feasibility of impact evaluation. Although there are nuances to this depending on the design, there are a number of common considerations around outcomes, which we consider in this section before turning to the specifics of each research design.

For both the RCTs or QEDs considered in the subsequent sections, the primary outcome would be a measure of paid employment. For DFNPS, it is particularly important that young people enter “sustainable employment” that is both paid and sustainable, defined as employment that is not casual and does not involve a zero-hours contract.

For Youth Futures, ensuring young people achieve “good work” is particularly important (Youth Futures, 2026). Additionally, it is intended for evaluations commissioned by Youth Futures to evaluate the impact of interventions on outcomes in a semi-standardised fashion to increase the rigour of and allow comparability between evaluations, utilising Youth Futures Outcomes Framework (forthcoming), although this will not be prescriptive and outcome selection will ultimately remain driven by the specifics of each evaluation.

At present, DFNPS collects post-programme, follow-up data through ongoing contact with programme participants through colleges and job coaches. These data include measures of employment status, types of employment, hours worked per week, and hourly wage, and are shared with DFNPS in an anonymised format.

For a full-scale impact evaluation, consistent access to accurate, high-coverage individual-level outcome data would be required. If collected directly by colleges and job coaches, and shared with DFNPS, this would necessitate the establishment of Data Sharing Agreements (DSAs) to enable secure transfers of identifiable data to the King’s evaluation team. Appendix G outlines other options for collecting primary outcome data, such as through the Longitudinal Educational Outcomes (LEO) dataset. Relying on the LEO dataset, however, could imply that evaluators are unable to access relevant data on time to produce results before the end of 2028, as there is a significant time lag in public administrative datasets.

Secondary outcomes could include self-confidence, social and communication skills, health and wellbeing, as well as financial independency, based on the ToC. These could be collected either via an endline survey with participants or less systematically assessed through an Implementation and Process Evaluation (IPE).

Across all proposed evaluation designs, a mobilisation phase would be required to finalise outcome measures and determine appropriate data collection time points. During the feasibility stage, it was proposed that outcome data be collected approximately nine months after completion of supported internships, allowing sufficient time for participants in the treatment group to realise potential employment-related impacts.

4.2. Randomised Controlled Trials: general considerations

In this section, we consider the general feasibility of an RCT in the context of DFNPS. RCTs offer a robust approach for identifying the impact of DFNPS on participants’

employment outcomes. The following sections consider two potential RCT designs, focusing on two different target cohorts:

- **Scale-up delivery to SEND students without EHCPs:** The first would involve expanding DFNPS provision to students with SEND, specifically those with a learning disability, with autism, or both, who do not have an EHCP, i.e., are not currently eligible for the programme due to a lack of funding availability. This could be within existing DFNPS colleges or new ones, but most would be expected to be colleges that DFNPS already have a relationship with. Individuals would be randomly assigned to DFNPS (treatment group) or BAU provision (control group).
- **Scale-up delivery to new colleges:** The second would involve expanding DFNPS provision to the current cohort (students with an EHCP, and with a learning disability, with autism, or both) within a new set of colleges in England. Individuals would be randomly assigned to DFNPS (treatment group) or BAU provision (control group).

Both designs would produce robust results and are practically feasible, provided certain conditions are met.

Crucially, scaling up delivery to either of the new populations overcomes potential ethical concerns, as it avoids restricting access to DFNPS for those individuals who would currently benefit from the programme. In either scenario, instead, DFNPS would be delivered to populations who currently could not participate in DFNPS.

This section will first discuss the common elements across both RCT designs, highlighting general enablers and barriers. Then, in the following subsections, the report looks specifically at the feasibility of each of the scale-up designs.

4.2.1. Methodology

In the RCTs, in line with current DFNPS delivery approaches, candidates would be referred by their colleges to attend site assessment days, with DFNPS and King's informing colleges about the referral criteria. While each RCT approach would have specific eligibility criteria to reach participants with or without EHCP, both RCTs at a minimum would have the following eligibility criteria. Participants must:

- Have a learning disability, autism, or both
- Be enrolled as students in a participating college
- Be between 16 and 24 years of age at the time of referral
- Have the Right to Work in the UK
- Not have previously participated in a supported internship programme

- Not experience severe mental health issues

Consent to participate in the RCT would be obtained during site assessment days (though there could also be a preliminary consent process prior to site assessment days). Those consenting to participate would undergo an eligibility check. The results of the eligibility check and the survey completed to collect baseline data would be connected to candidates' rubric scores using unique identifiers. Candidates who meet the eligibility requirements would then be randomised into either a treatment group, receiving an offer of a supported internship, or a control group, receiving an offer of BAU support through their college.

Outcome data would then be collected again following completion of the internship (endline). There could also be interim data collection points (midline). Outcome data could be collected through a survey or by accessing colleges' records and administrative datasets. Appendix G summarises the different options explored to access outcomes data. So far, the most practical approach for the primary outcome data collection involves receiving data directly from colleges through individual DSAs between colleges and King's; however, a survey with participants might be needed to explore secondary outcomes. A mobilisation phase prior to the trial should assess the completeness and consistency of the records across colleges and consider any necessary adjustments based on evaluation priorities, funding, timelines for data gathering, and the potential use of other sources from LAs, if relevant.

Data on outcomes, connected to participants' initial survey, their rubric sheet scores, and treatment allocation, would then be analysed to capture the key characteristics of the trial sample, including describing participants' sociodemographic characteristics, describing the distribution of outcome variables, and carrying out randomisation balance checks to ascertain whether treatment and control groups are comparable. Outcomes for treatment and control group would then be compared to identify the causal impact of DFNPS on employment outcomes, and available covariates could be introduced to increase the overall power of the trial to detect a treatment effect.

4.2.2. Enablers

Overall, both RCT designs would be robust approaches to generate high-quality causal evidence on the effectiveness of the DFNPS programme as an intervention in improving employment outcomes for young people with a learning disability, with autism, or both, and the programme is well suited to be evaluated through an RCT.

First, evidence from site assessment days observed in the 2025 cohort study suggests that DFNPS is already oversubscribed, creating opportunities to randomise participants.

Second, the eligibility criteria are likely to be easy to implement and apply for an RCT, as robust eligibility checks are already embedded in DFNPS' standard processes, and could easily be adjusted to meet the needs of an RCT. It also creates a clear point where randomisation can take place. Further, working with the existing eligibility criteria increases external validity. Because participant selection can be based on established eligibility criteria that are already used in routine delivery (or could be adjusted minimally for the purposes of the trial), the trial population would closely reflect a wider relevant population of young people.

Third, in both cases, as discussed above, scaled-up provision ensures an RCT can be delivered ethically. This is because it allows to create a control group who would have not been given the chance to access the programme outside the trial. This also allows to mitigate potential risks of contamination across groups.

Fourth, as we discuss in further detail in the next subsections, RCT designs are accepted and endorsed as a practically feasible evaluation approach by Youth Futures, the DFNPS leadership, provided certain delivery conditions are met.

Fifth, the rubric sheet and other research instruments piloted in the feasibility were widely, consistently, and successfully used across site assessment days. While researchers identified a number of inconsistencies, mainly around data recording, the final records included in the feasibility study were of a high quality, showing that future full-scale trials could use the same instruments successfully to evaluate the impact of DFNPS.

Sixth, we anticipate that the plausible eligible effect size of this intervention will be large, which means we can produce a well-powered trial with a relatively small sample size. Given the complexity of standing up provision with new partners, the relatively low required sample size required is a point in favour of these designs.

Seventh, and final, a prospective design allows us to conduct primary data collection with outcomes and data collection tools tailored to the intervention and cohort – whereas a retrospective evaluation would be constrained to what had already been collected.

For the EHCP cohort, one potential avenue to measure outcomes through administrative datasets is accessing National Client Caseload Information System (NCCIS) data collected by LAs. This dataset covers young people aged 18 to 24 if they have an EHCP (Department for Education, 2025). NCCIS data provides supporting information on participation and transition patterns, such as from Not in Education Employment or Training (NEET) to employment. Further work would be needed to assess how LAs collect this information, its reliability and consistency, and whether there is room to extend collection to include participants up to 25 years old. This dataset though does not

collect information from young people in these age groups who don't hold an EHCP, which therefore limits its usability for evaluating effects in such cohort.

4.2.3. Barriers

Setting up the trial

The key challenge, which would need to be further considered during a mobilisation and pilot stage, is to prepare DFNPS and its stakeholders, such as colleges and host businesses to deliver an RCT alongside usual provision. This requires ensuring enough combinations of colleges and host businesses are recruited to reach adequate sample sizes, while also securing buy-in to the trial design. To support this, the evaluator could allocate resources including by running training and information session with colleges to explain how referrals happen, how participants are randomised into treatment and control groups, and the provision for BAU support dedicated to those in the control group. While the trial would be delivered based on existing DFNPS approaches and research instruments already piloted during the feasibility study, a mobilisation phase would still be necessary to align objectives across all stakeholders in the trial, and examine what specific incentives can guarantee strong participation, both at recruitment and follow up.

Operational capacity

Running the RCT alongside usual delivery would also require DFNPS, Youth Futures, and King's to work together to ensure that they can meet the costs, both in terms of time and financial resources, required to run a trial. The trial's operational feasibility will depend on DFNPS' operational capacity to deliver an RCT at scale. The feasibility study recorded instances during some selection days where associates struggled to manage large numbers of candidates, highlighting the need to ensure sufficient staffing, resources, and processes are in place to support recruitment, delivery, and evaluation activities. Delivering the RCT would also require the creation of new supported internships for the new cohort of participants. These placements could be developed through existing college partnerships, or by engaging new colleges in different areas where existing host businesses are also located.

Ethical considerations

Alongside these delivery considerations, ethical safeguards will also be required to ensure the trial can be delivered without affecting existing provision. A key challenge will be establishing a clear distinction between the existing DFNPS cohort and the evaluation

cohort (non-EHCP participants or participants in new colleges). This will be important to avoid unintended impacts on current participants or ethical concerns around access to the programme. In particular, colleges would need to understand and accept the implications of randomisation, including how participants are allocated to treatment and control groups.

Data collection and attrition

Embedding an RCT on DFNPS' usual provision also poses data challenges. Systems will need to be put in place to ensure that participant records are registered consistently, especially if other evaluation designs, such as an RDD, are implemented in parallel with the 2026 cohort.

Moreover, the risk of attrition and missing data consequently is considerably high if the design relies on survey data. This risk applies across both treatment and control groups, as maintaining contact for midline and endline data collection with participants beyond college could prove challenging once students leave their institutions. High attrition could reduce statistical power. Therefore, during mobilisation, it would be key to explore how (in addition to/ instead of primary data collection) evaluators can obtain follow-up outcomes for the treatment and control group, including colleges records or administrative datasets.

While survey-based primary data collection with both treatment and control participants would be necessary to capture a range of secondary and intermediate outcomes – such as social skills, wellbeing, health and financial independence – it presents challenges in maintaining engagement with participants over time. This is particularly relevant in an RCT context, where outcome data need to be collected prospectively from both treatment and control groups. To mitigate risks of attrition, mobilisation could explore different approaches to maintaining participant engagement and collecting follow-up outcomes. For instance, a US DFNPS trial successfully adopted a 3-question phone interview to capture employment outcomes.

Allowing for measurement of these secondary outcomes is one of the crucial advantages of RCTs over QED in general. Therefore, it would be important to maintain engagement across both treatment and control participants, for instance through effective use of incentives tied to outcome data collection. Administrative datasets and college records could, however, provide a useful source for measuring primary employment outcomes, reducing reliance on survey data for these measures.

Despite this, college records and administrative datasets present different challenges and would need to be considered separately. For college records, further work would be needed to establish whether colleges collect relevant outcome information, the

consistency and completeness of these records, and whether data can be shared in a format suitable for evaluation purposes. To mitigate this risk, it will be essential to secure the commitment of colleges to collect follow-up data, ideally using unique student identifiers to facilitate linkage to available administrative data.

In contrast, accessing wider administrative datasets would depend on the availability and timing of data linkage processes, as well as the ability to match trial participants to relevant records. Initial conversations with colleges show there is a risk that evaluators wouldn't be able to collect follow-up outcomes data when students leave college. The feasibility study explored the feasibility of using several administrative data sources to collect participant outcomes, including the Individualised Learner Records (ILR) and the LEO dataset. Regarding the ILR, it should be noted that the ILR historically included a "Destination" variable providing detailed information on post-learning employment and education in a standardised manner. However, this variable has been discontinued from the 2024/2025 cohort onwards, as the DfE has shifted towards reliance on linked administrative datasets (HMRC, DWP, HESA/UCAS) accessed via the LEO dataset (Camden, 2024).

Generally, administrative datasets would be useful for measuring primary outcomes relating to employment within an RCT, but regardless of what decisions are made during a mobilisation phase regarding administrative datasets, any evaluation would ideally explore a range of secondary and intermediate outcomes – such as social skills, wellbeing and health and financial independency – which would need to be collected through surveys. Allowing for measuring these secondary outcomes is one of the crucial advantages of RCTs over Quasi-experimental designs in general. This means it would be important to maintain engagement across both treatment and control participants, for instance through effective use of incentives tied to outcome data collection.

4.3. RCT with scale up to participants without EHCPs

This section explores the RCT involving scale-up to SEND students without EHCPs. It includes more information and considerations about the methodology, enablers, barriers, and required sample sizes. It also discusses the acceptability of the design and specific ethical considerations that should be considered before its implementation.

4.3.1. Practical feasibility and robustness

DFNPS's current provision focuses on young people with SEND with EHCPs. However, evidence from the feasibility study interviews with colleges, workshops with DFNPS, and

analysis of official statistics suggest that DFNPS could benefit a much larger number of students with SEND, irrespective of whether they hold an EHCP or not. This is because there are SEND students without EHCPs that arguably face similar barriers in education and their transitions to employment than those with an EHCP. The key distinction between those with and without an EHCP is that the former have gone through a complex and challenging administrative process in order to receive additional support and resources beyond what colleges can typically provide, which might indicate greater barriers that the young people face in education and their transition to employment, although this is not a perfect indicator; as previously explained, obtaining an EHCP requires families to navigate complex administrative systems which might put some off from applying, and delays in the issuing processes are frequent, meaning that there are young people that, despite qualifying for one, do not have an EHCP.

Importantly, EHCP-holders are a smaller proportion of the larger population of students with SEND. According to recent figures for England, 19.5% of all pupils have special educational needs. This breaks down to 14.2% who receive SEND support without an EHCP, and a further 5.3% who hold an EHCP (DfE, 2025b). Interviews with three college representatives during the feasibility study also confirmed that SEND learners without an EHCP composed a large group of learners within their colleges⁵. For instance, two colleges reported having 400 and 800 SEND learners, respectively. As such, these figures suggest that the potential pool of suitable DFNPS participants extends beyond EHCP-holders, with a substantial proportion of young people with SEND not holding an EHCP.

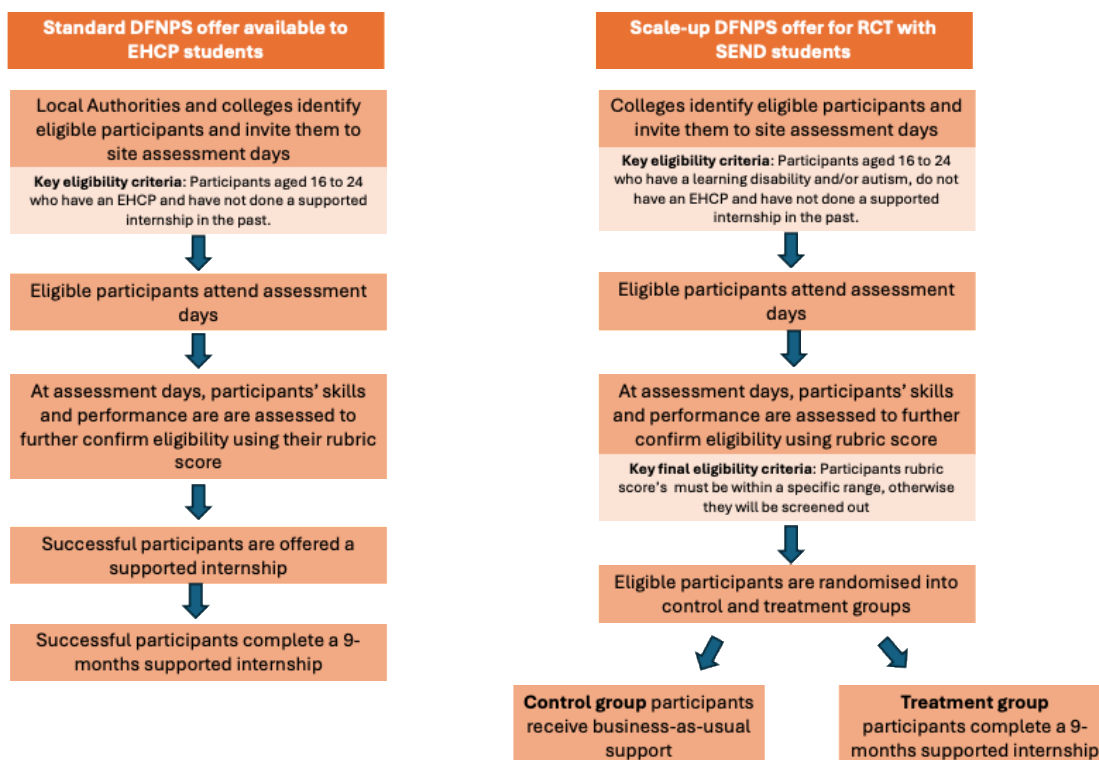
Crucially, this scaled-up approach would provide a valuable opportunity to evaluate DFNPS among a population that is not currently eligible for the programme. This would generate new evidence on programme effectiveness for the population of SEND students without EHCP, which would prove to be a valuable insight that could support the scale up of funding to other cohorts, for instance through relaxing the EHCP eligibility requirement for participants in the programme beyond the trial. The trial would also contribute to the wider evidence base on supported internships for a broader population of young people with SEND, rather than just those with EHCPs.

Additionally, the findings may also be informative for DFNPS' existing target population, provided that SEND students with and without EHCPs are sufficiently comparable in terms of their characteristics and barriers to employment. If this assumption holds, the results would have broader applicability to the current DFNPS model. If, however, important differences exist between the two groups, the trial would still provide robust

⁵ It is worth noting, that it was only possible to validate this with two colleges, given the timeframes for this project. This figure is a guidance point for recruitment, as this pool will likely shrink later during eligibility checks and when obtaining consent for the research.

evidence on the effectiveness of DFNPS for a new population. The enablers section, below, reflects further on whether the comparability assumption can be met robustly. Overall, this feasibility study finds that this approach is robust and feasible, and could capture DFNPS' effect on participants. Figure 7, below provides an overview of how this method would work, compared to the current DFNPS offer.

Figure 6: RCT approach comparison with scale-up to SEND students without EHCPs



Enablers and barriers

This section outlines the key enablers and barriers influencing the implementation of this methodology, providing further insight into its robustness and practical feasibility.

Referrals and eligibility criteria

One of the key requirements to deliver this RCT is to engage colleges to revise their referral approach, communicating both the change in target group, and making sure a larger number of SEND students without EHCP are referred to DFNPS. The change in target group would require a clear working definition of SEND for the trial and

collaboration with SEND officers within colleges to understand their current outreach practices.

Researchers should work closely with SEND officers to identify clear eligibility criteria, recognising that SEND identification methods vary across colleges. During the feasibility study interviews, one college representative mentioned that although there is a coding system to collect data when students join a college, there might be registration errors when it comes to disability types.

Currently, the DFNPS referral process involves a bespoke selection conducted by college tutors who identify students they believe to be eligible and who would benefit most from DFNPS. As observed during the feasibility study, this translates into most young people attending site assessment days being deemed eligible and suitable for the programme and being offered a supported internship. However, an RCT would require that a larger number of young people are referred to the site selection days so that a control group can be created. As such, early engagement with colleges in a mobilisation phase would be essential to source encourage a larger number of referrals.

Beyond adapting the referral process, a mobilisation phase should also consider how eligibility will be assessed for participants who do not hold an EHCP. Under current delivery, an EHCP provides evidence that a young person has SEND and requires additional support. For the proposed trial, alternative processes would therefore be needed to confirm that participants have the relevant SEND characteristics (e.g. a learning disability, autism, or both) and are suitable candidates for a supported internship. While the overall eligibility criteria outlined in the previous section provide a starting point, King's, DFNPS and Youth Futures would need to identify those who are best placed to undertake a supported internship, based on evidence on the characteristics of the new population of interest.

Connectedly, it would also be important to consider whether the new population would also shift participants' support needs. In the original population of EHCP holders, these needs would be assessed as part of obtaining the EHCP itself, and a similar assessment would have to be conducted either within colleges SEND support frameworks, or as part of the trial itself.

Comparing SEND students with and without EHCPs

Key to the generalisability of this trial design is to what extent SEND students with and without EHCPs are comparable. Feasibility study interviews with college representatives confirmed that both groups of young people share many similar characteristics, with research instead pointing mainly to differences in parents' characteristics like their knowledge of the SEND system (Adams et al., 2017) and to the importance of

socioeconomic status (Misik and Barnhardt, 2024) in determining whether an individual has an EHCP or not.

In an interview, one college representative echoed this evidence, sharing that the decision to pursue an EHCP often depends on parental or guardian engagement in what can be a lengthy process – one that not all families choose to undertake. Another interviewed college representative observed that students with SEND without an EHCP may face fewer barriers to entering the labour market, although they still experience many of the same challenges as those with an EHCP, raising questions on the potential need for eligibility criteria to exclude those who potentially obtain rubric scores that are too high in assessment days. A further step to increase the generalisability of the trial findings considering students with and without EHCPs could include a systematic review of existing evidence on the way these groups differ, to identify differing factors and capture them using participants' surveys during the trial.

Generally, a mobilisation phase would be an appropriate moment to further investigate the employment outcomes of SEND students without EHCPs, and how these compare to the outcomes of EHCP-holders. If students without EHCPs face lower barriers to accessing the job market, we would expect the effect sizes of receiving DFNPS as a treatment to be smaller than what would be measured in a sample of EHCP-holders, as measured for example in the US trials of Project SEARCH.

At the time of writing this report, EHCP-eligibility is currently the object of political debate and potentially the object of reform, as suggested in comments from the former Schools Minister Catherine McKinnell (Adu, 2025). No concrete plans have been announced, so any policy changes affecting the 2026 cohort are unlikely. Any shift in EHCP eligibility would affect the pool of participants to a DFNPS trial and to DFNPS's current provision.

On balance, evidence still shows that the trial would not target a completely new population for DFNPS, and findings from the feasibility study are likely to hold broadly true for the trial target group as well. This likely means that the proposed trial would be a robust approach to evidence the impact of DFNPS, both for young people with SEND without an EHCP, but this could likely also be generalised to their normal cohort of EHCP learners.

Rubric sheet and other research instruments adjustments

Beyond eligibility criteria and referrals, scale-up to SEND students without EHCPs would require considering whether adjustments are needed to the rubric sheet and its use during site assessment days. First, it would be crucial to ensure it works well to assess SEND students without EHCPs. Second, it would also be crucial to consider whether a threshold rubric score, or minima and maxima rubric scores, are required as part of the eligibility

requirements to take part in the trial and DFNPS. In the feasibility study, candidates who accepted supported internships had rubric scores between 55 and 98, with 87% scoring over 70. For the RCT, King's and DFNPS would need to agree on an appropriate value for the threshold for the new group to be eligible, and whether any candidates scoring too high or too low would also be deemed ineligible.

A further element to consider is that should substantial differences arise on the way rubric sheets are designed and used, DFNPS would have to ensure that the approaches used in their normal and trial provision are not mixed up, for instance in a college that is involved in both normal and trial provision. DFNPS has expressed a preference to reduce this risk by holding separate site assessment days exclusively for trial participants.

It would also be important to validate existing research instruments as used in the feasibility study, such as the referral, consent, and outcome survey instruments. This would be done through conducting cognitive testing with a group of SEND learners without an EHCP to validate whether the questionnaire is suitable for this population. It would nevertheless be reasonable to assume that surveys, consent forms, and information sheets, which have worked well with EHCP learners, will also be suitable for the cohort of SEND learners without EHCPs.

BAU provision

The proposed scale-up approach could also have consequences for the provision of BAU support to participants randomised to the control group. Feasibility study interviews with colleges indicate that, in the absence of a DFNPS placement, young people with EHCPs may be offered a variety of alternative provisions, such as supported internships from different providers (e.g., charities such as Mencap or the Shaw Trust), short courses, or vocational training. Most supported internships, however, rely on EHCP-linked funding, meaning eligibility is often restricted to those with statutory plans. Despite this, the feasibility study showed the offer can vary on a college-by-college basis, and by colleges' relationships with LAs. For instance, while one college interviewed reported having no other alternative supported internships, another received council funding to offer supported internships to SEND students without EHCPs. Overall, while supported internships are not typically available for students without EHCPs, there are other pathways for SEND learners such as apprenticeships and work experience placements, or other SEND-specific programmes provided by colleges or LAs that constitute BAU. Nevertheless, we would expect BAU to be generally weaker for the SEND students without EHCPs than the normal EHCP population.

Understanding these variations will be crucial to interpreting the trial results accurately and to understand how it relates to standard DFNPS provision. Some of these alternative internship programmes or other types of support could positively influence employment

outcomes, narrowing the difference between the treatment and control group. If the BAU for the non-EHCP cohort is indeed weaker for selected colleges in the trial, the same RCT is likely to yield a larger effect size for the non-EHCP cohort than for the cohort with EHCPs, assuming groups are otherwise comparable. Consequently, it is important to examine the BAU in detail in the full-scale evaluation to understand how it compares with the standard DFNPS BAU. More generally, as there are plausible arguments for the effect size among non-EHCP pupils to be either larger or smaller, it is recommended to maintain large sample sizes (see following section) to ensure sufficient power to detect a statistically significant effect of the DFNPS programme.

Overall, the feasibility study, supported by ongoing conversation with colleges and DFNPS partners, indicates that addressing these challenges to increase the robustness of an RCT trial of DFNPS would be feasible as part of a trial mobilisation phase. It is therefore concluded that this design is feasible and capable of producing robust evidence on the causal impact of DFNPS on employment outcomes.

Sample size

Table 10 shows the indicative power calculations for an RCT of DFNPS with SEND students without an EHCP⁶. These calculations estimate the required sample size to capture various Minimum Detectable Effect (MDE). The MDE refers to the smallest effect that an intervention must produce in order to be detected as a statistically significant difference on the primary employment outcome between the treatment and control groups. These are then translated into risk ratios as Minimum Detectable Effect Sizes (MDESs).

Sample sizes were estimated for three MDES scenarios:

1. An MDE of **56 percentage points** as this is the effect detected for Project SEARCH in the US trials,
2. An MDE of **37 percentage points** that corresponds to a break-even cost–benefit scenario for an estimated cost of programme delivery of £12,000, one possible cost figure estimate provided by Youth Futures,
3. An MDE of **18 percentage points** that represents a smaller effect scenario to provide a comprehensive assessment across a range of effect sizes.

A few parameters are fixed across all calculations:

- Given the absence of baseline employment data for this cohort, the power of covariates is assumed to be low (0.10).

⁶ ICCs are not included in the table, as they are not relevant in this design.

- All calculations assume a 95% confidence in the estimations which corresponds to an Alpha of 5%, and a statistical power of 80%.
- All calculations assume a two-sided distribution, and an even allocation of individuals between the treatment and control group.
- Randomisation is conducted at individual level.
- The employment rate is 10% in control group, based on current employment figures for the target cohort.

In addition, break-even costs are estimated for each of the MDES scenarios. These estimates consider a five-year horizon and a diminishing persistence profile for benefits, as computed through the economic evaluation presented in Appendix H.

Table 10: Power calculations for an RCT with SEND students without ECHPs

Parameters	Scenario 1	Scenario 2	Scenario 3
MDE (percentage point change)	56	37	18
MDES (risk ratios, control group employment rate: 10%)	6.6	4.7	2.8
Break-even cost of programme delivery	£18,162	£12,000	£5,838
Total number of participants required across treatment and control (complete cases)	30	32	120

Attrition and definition of complete cases

All estimated sample sizes refer to complete cases, defined as participants who complete the internship and all waves of data collection. Attrition may occur post-randomisation (for example, non-response at midline and endline surveys).

Survey attrition is a particularly significant risk when working with young people. The mobilisation phase should develop strategies to maximise engagement and response rates, especially among control group participants. For evaluation purposes, individuals who are randomised but subsequently withdraw from their supported internships should still be included in the analysis, in line with an intention-to-treat approach.

Sensitivity to key assumptions

The required sample sizes are sensitive to changes in modelling assumptions. For example, if a one-sided test is used instead of two-sided test, required sample size decrease because the analysis rules out the possibility of negative impacts. Under a one-sided distribution, the number of participants to capture the Scenario 3 MDES would

fall from 120 to 94. The assumption behind a one-sided distribution is that the effect would be positive, which aligns with the US trial which found large positive effects.

Sample size requirements are also sensitive to the assumed power of covariates. If demographic data could be obtained from colleges, which could increase covariate power e.g., from 0.10 to 0.20, the required sample size for Scenario 3 would fall from 120 to 114.

Recommended sample size

Taking these considerations into account, this study recommends recruiting a total of 150-200 individuals (75-100 per trial arm) to achieve 120 complete cases. This recruitment target is intended to ensure a sufficient number of referrals after accounting for attrition, which may take several forms, including programme dropout, research withdrawal, and survey non-response, and is indicatively expected to range from 20% to 40%. This could be achieved by engaging approximately 10 colleges, each referring around 20 candidates to site selection days, allowing for attrition both pre- and post-randomisation. The mobilisation phase would need to provide more exact estimates for each college, based on how many places each of the host businesses have for supported internships, and more precise estimates for what attrition would be at the different stages; that is, how many referrals to site assessment days would be required to achieve 120 complete cases. Another consideration will be logistical constraints for DFNPS and financial constraints for Youth Futures funding the trial.

The target of referring 150-200 individuals appears feasible given the size of the eligible population. DfE data indicate that 1.3 million students in England have SEND, representing 14% of the whole student population (GOV.UK, 2025). Among those young people with SEND, 58% will not have an EHCP (DfE, 2026a; DfE, 2026b).

Based on prior trial experience, it should be noted that a study with fewer than 50 complete cases is not recommended, regardless of whether the MDES is considered appropriate. Such small samples substantially increase the likelihood of randomisation failure, as the law of large numbers doesn't apply, and so there is the potential that, by chance, the randomisation results in an unbalanced sample. Such imbalances can result in biased impact estimates if the differences in outcomes reflect underlying group differences, rather than the effect of the intervention.

Given the large population of young people with SEND without an EHCP, even within individual colleges, the calculations suggest that a well-powered design could be achieved by engaging a relatively small number of larger colleges. However, this approach would limit the external validity of the results, particularly in terms of geographical diversity and population characteristics. It might also not be feasible for

DFNPS who need to find host businesses in the local area. During the mobilisation phase, these trade-offs should be carefully assessed to balance statistical power, external validity, and financial and practical implications of delivering the programme.

Finally, while Scenario 1 reflects the effect size observed in the US Project SEARCH trial, there are important contextual differences. The US studies focused specifically on young people with Autism Spectrum Disorder, a population with lower employment rates and likely to face higher barriers to employment. The proposed UK trial would also include learners with a learning disability. As a result, the expected effect size may be smaller, as interventions often generate larger impacts for groups facing greater barriers to employment.

4.3.2. Acceptability

The possibility of expanding the provision to SEND students without EHCPs was initially suggested by DFNPS during the feasibility study's impact workshop and the approach has enjoyed continuous support in subsequent conversations. The proposed approach was also presented to DFNPS staff at mid- and senior-levels, including their CEO and their board. They expressed overall support for the design, providing that the trial can be structured in a way that doesn't negatively affect current provision for young people with EHCPs.

4.3.3. Working relationships

Explaining the RCT approach

A key requirement for delivering the evaluation would be to clearly explain the RCT approach to stakeholders, especially SEND officers who are responsible for referring and supporting students, and colleges that run site assessment days in collaboration with host businesses and DFNPS. These assessment days will be the primary sources of key trial data.

During a trial mobilisation phase, training and information sessions would be required to explain the purpose of the trial, how it will operate in practice, and the roles and responsibilities of each stakeholder involved in delivery. This will be essential to ensure consistency of implementation and buy-in across all partners.

Engaging and scaling host businesses

Scaling up DFNPS to include SEND students without EHCPs would require engaging new host businesses capable of delivering supported internships and site selection days. This could be achieved either by expanding provision with existing college partners, or by

identifying new colleges through which businesses in different geographies and communities can be reached.

DFNPS already plans to extend its provision for the 2026 cohort, with 46 new programmes involving new host businesses. This planned expansion suggests that sufficient supported internship places could be created to support the RCT. DFNPS has indicated that these new programmes could be dedicated solely to SEND learners without EHCPs, suggesting that there are no significant barriers to scaling up delivery from the perspective of working with enough host businesses.

Identifying and partnering with colleges

Suitable colleges will also need to be identified and paired with host businesses. DFNPS had established relationships with at least 85 colleges in 2025 and has historically worked with approximately 150 colleges in England. These existing relationships can be leveraged for recruitment into the trial. Several colleges have already indicated willingness to refer young people to new programmes and support the delivery of supported internships for SEND people without EHCPs.

Where additional colleges are required, DFNPS' existing networks and operational expertise could be leveraged. However, recruitment and onboarding of new colleges might affect timelines. Close monitoring during the mobilisation phase would be important to identify emerging pressure points and manage potential risks early.

Evidence of demand from colleges

Feasibility study interviews suggest there is demand from colleges for supported internship for SEND students without EHCPs. Representatives from two colleges expressed strong interest in programmes that would offer new progression options for this group. A third college reported that it currently runs a supported internship programme for SEND students without EHCPs, but on a very small scale, with less than 10 supported internships currently being offered, suggesting that this has been identified as an area of provision that colleges would want to cover.

Data collection with colleges

Although many relationships with colleges and businesses are already in place, or will soon be established, a mobilisation phase would be crucial to secure organisational support and commitment to collecting new data from participants and sharing existing data required for the trial. This will involve setting up data sharing agreements, and especially for any new colleges, establishing new data collection processes from scratch which may be time- and cost-intensive.

SEND officers

Finally, researchers would have to work closely with SEND officers to refer potential participants from a broader population than DFNPS has previously engaged. This would place additional pressures on SEND officers, who already face capacity and resource constraints, potentially affecting referral timelines and their capacity to support SEND students. Early insights from a mobilisation phase should inform practical risk mitigation strategies to ensure the trial remains feasible and on schedule.

4.3.4. Ethics

As discussed before, scaling-up provision allows for an ethically sound approach to deliver a RCT of DFNPS, as it avoids withholding existing support from eligible learners. Nonetheless, several other ethical challenges would have to be considered and addressed.

Defining eligibility of SEND without EHCP students

Unlike SEND students with EHCPs, who are assessed using a legal framework to receive their certificate, SEND status for students without EHCPs is typically determined by schools' and colleges' internal assessments. These practices can vary considerably across institutions, creating a risk of inconsistency in how eligibility is defined and applied, and undermine perceptions of fairness and transparency in the trial. To mitigate this risk, during the mobilisation phase, evaluators would have to work closely with DFNPS to establish clear and standardised eligibility criteria for inclusion in the trial. These criteria would be communicated consistently to colleges. They could, for example, be based on colleges' SEND registers or on institutional assessments aligned with the SEND Code of Practice, especially with regards to the provision around broad areas of need⁷.

Variation in BAU provision

Another ethical consideration relates to the variation in BAU support available to control group participants. Feasibility study interviews indicated that provision differs substantially across colleges. In some institutions, supported internships are not offered to students without EHCPs, meaning that participation in the RCT would not reduce existing opportunities for control group learners. In other colleges, however, supported internships are available to SEND students regardless of EHCP status, which could result in

⁷ SEND code of practice: 0 to 25 years. (2024, September 12). GOV.UK. Available [here](#).

differing levels of access between treatment and control groups. It would therefore be important to manage expectations among both colleges and learners during mobilisation. To mitigate any perceived unfairness, evaluators would have to ensure that all students and colleges are fully informed about the rationale of the trial and the randomisation process. Participants allocated to the control group would also be made aware of alternative opportunities available to them, such as other employability support offered through their college or by DFNPS.

4.3.5. Feasibility assessment summary

The overall assessment of this methodology is outlined in Table 11: Feasibility assessment summary below. The feasibility study found that the proposed RCT design is methodologically feasible, as DFNPS' existing referral and assessment processes can readily support randomisation. Feasibility is strengthened by the presence of a large and broadly comparable population of SEND learners without EHCPs, among other enablers.

The findings indicate that a sufficiently large sample can be recruited to achieve adequate statistical power, particularly in the light of DFNPS's planned expansion. Overall, this design is acceptable to DFNPS leadership, who has expressed support for extending provision, provided that their current cohort of learners remains unaffected.

Successful delivery will depend on establishing and maintaining strong working relationships with colleges and new host businesses, with the overall feasibility depending on effectively leveraging DFNPS' operational expertise to secure these partnerships. Finally, the trial is considered ethically sound, as it expands access to supported internships rather than withholding opportunities from DFNPS' current target group.

Table 11: Feasibility assessment summary

Approach	Robustness	Practical feasibility	Acceptability	Working relationships	Ethics
RCT with scale up to SEND young people without EHCP					

4.4. RCT with scale up to new colleges

In this design, the impact evaluation of DFNPS would be delivered by scaling up provision to randomly assigned new colleges that are not currently offering DFNPS to

EHCP students, comparing outcomes of students at these colleges to those at colleges randomly assigned to not start offering DFNPS.

This would allow both DFNPS and the trial to focus on the same target population currently served by DFNPS, allowing for the impact evaluation to use the current approaches to recruit, screen, and support participants. Conducting the RCT with new colleges not currently associated with DFNPS would avoid affecting provision in existing DFNPS colleges, who would not be part of the trial.

However, this approach would raise significant challenges when it comes to time and financial resources to engage new colleges, familiarising them with DFNPS procedures, policies, and approaches, and also to identify new supported internship opportunities with DFNPS business partners.

4.4.1. Practical feasibility and robustness

Enablers and barriers

Reaching out to new colleges

To deliver this scale-up RCT, a mobilisation phase would be required to identify and include enough new colleges interested in offering supported internship to their students through DFNPS. With DFNPS provision being already so widespread across the country, identifying the new colleges could prove difficult, especially as those not participating in the programme could have material reasons or hard-to-change circumstances for not doing so. Beyond identifying and including the new colleges, this scale-up approach would also need to consider the cost and challenge of training college staff on the DFNPS referral and delivery processes, and on the trial processes, including participants' randomisation, eligibility screening, and outcome measurements. It is especially worth highlighting the risk that recruitment and RCT familiarisation activities, combined, could stretch the RCT delivery timeline beyond the time-horizon Youth Futures would be able to fund the study.

Identifying new businesses to provide new supported internships

Additionally, to accommodate the increased demand of the scale-up trial, new businesses would also have to be recruited, or new opportunities identified with existing partnered businesses to offer additional supported internship opportunities for the new colleges, without impacting existing provision. A feasibility study workshop with members of DFNPS highlighted that there is interest among new host businesses to offer supported internships, which would facilitate the scale-up of provision.

DFNPS and trial processes consistency

For this RCT, DFNPS could benefit from having a consistent approach, broadly unchanged from their usual provision, to target both their existing colleges, and new ones included to deliver this RCT. This would simplify the DFNPS programme delivery processes, and limit concerns from potential issues arising from a failure to use different approaches across colleges in the trial and colleges already receiving DFNPS provision.

In sum, while methodologically feasible and robust, this design would raise substantial concerns around the capacity of identifying a large enough sample for estimating causal effects, together with concerns on how it could be delivered within funding timelines, and without incurring excessive delivery costs.

Sample size

While slightly different target groups – EHCP learners vs SEND learners without EHCPs – the power calculations and the required sample sizes proposed in Section 4.3 remain valid for both RCTs.

Some differences are, however, worth highlighting. It is likely that DFNPS will need to secure a larger number of colleges than the previous RCT, as there are fewer EHCP students compared to SEND without EHCP students in each college. As part of the feasibility study, 412 candidates were reached and completed the survey based on the involvement of 85 colleges. Based on these numbers alone, an RCT would need to work with around 30-40 new colleges (out of the around 212 FE colleges in England) to reach the 150-200 suggested referrals to site assessment days. It may be that colleges could be encouraged to refer more students with EHCPs each, but the total pool of potential eligible students is nevertheless smaller than in the SEND without EHCP trial.

As already discussed, there are several challenges with involving a larger number of colleges, including DFNPS' ability to recruit those in a short time period and it will also make the overall trial more expensive, as funding is expected to cover staff time per college for the DFNPS model to operate effectively, including dedicated tutors and job coaches who are normally employed through the college.

4.4.2. Acceptability

During the feasibility study workshop with DFNPS, delivery staff have signalled that the acceptability of this design is lowered by the potential costs involved in reaching out to new colleges, the longer timelines required to deliver this, and the higher risk to costs, timelines and programme fidelity associated with training a whole new group of college

staff to deliver DFNPS and the RCT. The acceptability is therefore perceived as lower than the alternative scale-up to SEND students without EHCPs.

4.4.3. Working relationships

While DFNPS have reported they have received interest from host businesses to provide further supported internships, a future scale-up RCT would still depend on whether they can recruit a significant number of new colleges. DFNPS' operational expertise and networks would be crucial for this process. A trial mobilisation phase would need to last for a longer time than the mobilisation phase for the SEND with EHCP RCT due to the considerable efforts associated with recruiting enough colleges, develop those relationships, and provide information and secure commitment on trial processes, data collections, data sharing agreements, and so on.

4.4.4. Ethics

While the scale-up delivery allows for an ethical delivery of the trial, where no existing beneficiaries are withheld from taking part, this design would still need to consider ethical considerations around running an RCT, including around participants' informed consent, data management and processing procedures, and randomisation, as well as how BAU would be delivered to those in the control group.

4.4.5. Feasibility assessment summary

Table 12, below, summarises the overall assessment of this design. This design is methodologically feasible and would produce robust causal evidence through individual-level randomisation within new colleges, without disrupting the existing DFNPS's current offer. The RCT could be delivered in an ethically sound way, since it does not withhold provision from young people who currently benefit from it, instead it expands opportunities. However, the approach depends on recruiting and successfully establishing strong working relationships with a relatively large number of new colleges within a short time-period in order to align with Youth Futures timelines for funding. As such, this design entails substantial practical risks. Conversations with DFNPS confirm that, while they see the approach as technically viable, there are substantial reservations about the practical deliverability within fixed timelines and resource demands of this scale-up model.

Table 12: Feasibility assessment summary

Approach	Robustness	Practical feasibility	Acceptability	Working relationships	Ethics
----------	------------	-----------------------	---------------	-----------------------	--------

Individual RCT with scale-up to new colleges					
---	--	--	--	--	--

4.5. Matched Difference-in-Differences at college level

The next impact evaluation design considered is a retrospective analysis of historical DFNPS delivery using a DiD design with college-level aggregated outcomes. This design is scoped only for the primary outcome: employment (or "sustainable" employment, if possible). This DiD design seeks to estimate the effect of DFNPS by comparing changes in employment outcomes over time between those who took part in DFNPS (treatment group) and those who did not (comparator group). Specifically, this approach would compare the changes in employment outcomes at college level. An individual-level DiD is not feasible as DFNPS participants are typically at the point of transition into the labour market and therefore lack pre-intervention employment outcomes. Looking at individual outcomes aggregated at college-level overcomes this barrier by providing a baseline measure for employment before the introduction of DFNPS within each college.

4.5.1. Practical feasibility and robustness

This section explores the matched college-level DiD approach in more depth, including more information and considerations about the methodology, enablers, barriers, and required sample sizes. Overall, it finds that the DiD design is feasible and robust, provided certain assumptions are met.

Methodology

Design and identification strategy

To carry out a matched college-level DiD evaluation, administrative data could be used to estimate trends in college-level employment outcomes over time (by aggregating the individual-level employment outcomes of students attending the college who hold an EHCP or have a disability). The analysis could then test whether these trends changed significantly for those colleges who participated in DFNPS compared to those that were not. The central identifying assumption is that, absent the intervention, treatment (i.e., participating in DFNPS) and comparison (i.e., not participating in DFNPS) colleges would have followed parallel trends in employment outcomes. Subject to this assumption and

others discussed below, this DiD approach could allow estimation of the causal impact of DFNPS on employment.

To identify the treatment group, DFNPS would provide a list of colleges they have worked with, including the start and end dates. The initial pool for the comparator group would comprise all colleges in England that have not participated in DFNPS-supported internships, either at all or up to that point. From this pool, coarsened exact matching (CEM⁸) could be used to select a subset of colleges that are most similar to the participating institutions. While matching is not strictly required for a DiD design, it can substantially strengthen the plausibility of the parallel trends assumption. This is because matching would allow us to identify, from the pool of colleges in England, those that are most similar to the colleges who engaged with DFNPS in terms of a relevant set of pre-defined characteristics. While it is not possible to rule out the influence of unobserved characteristics across colleges, matching does help to mitigate sources of bias hence strengthening the case for parallel trends across groups. Matching could be based on publicly available college-level information, such as size, gender composition, and other relevant characteristics. Additionally, pre-treatment employment trends could be incorporated to further improve comparability.

Outcome data sources and outcome measures

Outcome data and information required to construct the comparator group could be drawn from administrative sources, particularly the ILR or LEO (which also contains ILR as well as HMRC data). During the set-up phase for this evaluation, an assessment should be conducted to determine which dataset (or a combination of both) is preferable.

As previously mentioned in this report, the ILR historically included a “Destination” variable providing information on post-learning employment and education in a standardised format, yet this variable has been discontinued from the 2024/2025 cohort onwards.

With respect to LEO, while provider-level data should be available, there are limitations in terms of linking data particularly at individual-level. In particular, it is unclear whether the evaluation could gain access to information on supported internships and disability status due to risks of disclosure associated with small numbers, which may be relevant to young people with an EHCP, even at a de-identified level. Therefore, it would be

⁸ Researchers recognise the strength of this method compared with others, particularly Propensity Score Matching. For more information see: King, G., Nielsen, R., Coberley, C., Pope, J. E., & Wells, A. (2011). Comparative effectiveness of matching methods for causal inference [Unpublished manuscript]. Institute for Quantitative Social Science, Harvard University.

recommended to keep both data source options open and decide once these issues have been fully assessed in the set-up phase of the evaluation.

Enablers

Availability of large-scale administrative datasets

A key enabler for a DiD approach is the availability of large-scale administrative datasets, including the possibility of linking a DFNPS marker at the college level to extract data from the ILR or from LEO. These datasets make it possible to construct a longitudinal panel that captures the timing of each college's participation in DFNPS and associated employment trends for young people with a learning disability, with autism, or both, within these colleges, both before and after college participation.

The ILR data (also contained within LEO) holds detailed information on learners' disabilities and learning difficulties, which can be used to identify the population targeted by DFNPS and to aggregate their outcomes at the college level. First, this enables the construction of baseline and post-adoption measures of employment outcomes, for both treatment and comparator colleges, which is central to the identification strategy. Second, it permits outcomes to be defined specifically for the population of interest, thereby reducing potential masking effects arising from aggregation across heterogeneous learner groups, which could otherwise attenuate estimated impacts, as discussed in the previous section.

Additionally, given that this analysis will exclusively focus on young people with an EHCP, it could be possible to measure outcomes through accessing NCCIS data collected by the LAs for young people aged 18 to 24. This is subject to the data being complete, reliable and consistently recorded across LAs and over time, which should be inspected further if this design were to be implemented.

Consistency across time

More broadly, using administrative datasets provides outcome measures that are consistent across time, with low attrition and with well-documented methodological changes where relevant. This strengthens the credibility of the historical analysis and support the production of robust evidence.

Measuring meaningful and sustainable employment outcomes

Another enabler is the availability of granular information on employment and broader EET status. This facilitates the construction of outcome measures that capture not only employment entry but also dimensions of how meaningful and sustainable the employment is, which DFNPS highlighted as a priority of the model's intended outcomes. For example, the ILR "Destination and Progression" variable provides detailed

information on post-learning activities, including full-time and part-time employment, self-employment, voluntary work, and progression to further and higher education. In addition, LEO data offers historical records of employment from the HMRC data that could be used to construct outcomes across comparable post-leaving horizons (e.g., 6, 12, and 24 months). These data enable the construction of outcome measures related to employment duration and intensity.

Data coverage

Finally, DFNPS has worked with a large number of colleges throughout many years, which are covered in the administrative data. In addition, there is extensive coverage in both ILR and LEO data across FE colleges to ensure a sufficiently large comparator pool. A DiD design, combined with matching, requires access to an adequate set of comparison colleges with similar pre-intervention employment trends. This strengthens the parallel trends assumption and enhances statistical power.

Barriers

Despite the overall feasibility of a college-level DiD approach, the feasibility study identified a number of important barriers and limitations that would need to be addressed in a full-scale evaluation.

Analytical limitations

This option has several analytical limitations inherent to its design. Aggregating outcomes at the college level reduces variation both across participants and colleges, and masks heterogeneity and reduces the level of detail of the analysis. While matching helps account for observable differences between colleges, unobservable factors remain unaccounted for. For instance, the decision to work with DFNPS might reflect a general focus on SEND, or ability of local employers to host the supported internship might reflect local labour market conditions more suited to SEND employment.

Additionally, important programme information – such as rubric scores or baseline work-readiness – is not captured in this design, compared to other designs incorporating this data.

A further limitation is that treatment is assigned at the college level rather than the individual level. This implicitly assumes that all relevant students within a participating college are exposed to DFNPS, which is unlikely to hold in practice. As a result, estimated effects are likely to be attenuated, effectively masking true impacts, because partial exposure within treated colleges will bias effect size estimates downward.

Data availability and access to Destination and Progression variables

Another key barrier relates to data availability and access, particularly with respect to the Destination and Progression variable within ILR. While DfE documentation indicates that Destination and Progression records have been collected up to the 2024/25 academic year, inconsistencies were identified across published ILR data tables. For example, in the most recent version of the data tables, the Destination and Progression variable appears to be available only up to 2014/15, whereas earlier versions indicate availability up to 2024/25.

The colleges that the King's evaluation team spoke with confirmed that they continue providing this information, suggesting that the data exists. It should be said that the ILR also includes alternative employment-related variables that could be used, such as employment status, earnings, and sustained employment, but the Destination and Progression variable would be the most comprehensive measure of post-learning destinations.

Inconsistencies in the timing of employment recording

Another barrier concerns variations in the timing at which employment is recorded within the ILR. Employment status monitoring fields capture learners' circumstances at entry to further education, and while these fields can be updated subsequently to reflect changes, recording practices vary widely across providers. For example, some colleges may update employment data only within the two-month window post-course, while others may continue updating records until the end-of-year cut-off.

While some of this variation may occur randomly and therefore not bias estimates, there is a risk of systematic bias if colleges change their recording practices upon engagement with DFNPS, for example, in response to programme monitoring or reporting requirements. A full evaluation would therefore need to assess whether participation in DFNPS is associated with changes in data recording behaviour. This could be examined through further discussions with DFNPS and participating colleges, alongside a detailed review of licence agreements and reporting requirements, to determine whether such influences might exist and whether they differ from those affecting comparator colleges.

Changes in data collection methods over time

Periodic changes in data collection methods within both the ILR and LEO datasets represent an additional barrier. While such changes are usually documented transparently in methodological notes, they introduce discontinuities that may affect longitudinal comparability.

A final barrier to consider is the impact of Covid-19 on administrative datasets. During this period, several temporary measures were introduced across administrative datasets, including disruptions to or suspensions of data collection. While the inclusion of fixed effects can partially control for some of the bias introduced by this, the pandemic remains an important contextual factor that must be acknowledged as a limitation of this type of historical analysis.

Heterogeneity in treatment effects

Another limitation relates to heterogeneity in treatment effects across colleges. The evaluation sample would include colleges at different stages of DFNPS implementation, and with varying durations for outcomes to materialise following programme delivery. As a result, a standard DiD estimate would capture an average treatment effect that may mask meaningful variation across colleges and over time.

Therefore, a full-scale evaluation should consider weighting approaches, where information is available, to account for the number of colleges engaged each year. It would be important to transparently report what factors might influence estimates, including the direction in which effects may be over- or under-estimated, if possible.

Defining and accounting for BAU

It is important to consider the influence of what BAU looks like in a DiD context. In an RCT, BAU is typically more consistent and easier to characterise across a pre-defined control group. In contrast, a historical DiD approach must acknowledge BAU provision that is likely to have evolved over time and to vary substantially across the approximately 212 FE colleges in England.

This heterogeneity in BAU provision creates a risk that observed employment trends may be influenced by factors associated with the wider DFNPS delivery context, rather than by the programme itself. Although fixed effects and time-varying controls derived from administrative data may help address some of these concerns, the risk of unobserved confounding remains, as is inherent in quasi-experimental designs.

It will therefore be important to examine whether the data provide credible evidence of parallel pre-intervention trends, whether there are no confounding factors that can influence the treated colleges trajectory other than DFNPS, and whether there is any indication that spillover effects from treated colleges have affected the comparison group.

Provided these assumptions, alongside the other requirements of a robust difference-in-differences design, are sufficiently supported, it will also be important to understand how BAU provision is reflected in the data. This will determine whether the findings should be

interpreted as the causal differential effect of DFNPS relative to BAU provision, or as the causal effect of the programme relative to no comparable provision.

Sample size

Indicative power calculations were conducted to assess the number of colleges required to detect a range of MDEs in a college-level DiD design. Statistical power in this context is determined by several factors: the number of colleges included in the analysis, the number of pre- and post-intervention time periods observed, and the variance of the outcome prior to treatment. The main outcome of interest is employment status.

Indicative sample size requirements were calculated for three MDE scenarios. These include:

- A large effect informed by the impact detected in the US randomised trial of Project SEARCH (approximately a 56 percentage point difference);
- A mid-range effect of 30 percentage points, informed by evidence from economic evaluations;
- A lower-bound effect of 10 percentage points, reflecting a more conservative scenario.

Baseline incidence rates and outcome variance

The variance of the outcome should be estimated based on its incidence rate in the target population. For the baseline employment rate, we use an incidence of 5.9%, corresponding to the reported share of adults aged 18 or over with LD who are in employment (Inside Government, 2025). If this design is taken forward to a full-scale evaluation, it will be important to refine this incidence rate for the specific population of interest, namely young people with a learning disability, with autism, or both, participating in further education. Ongoing work by Youth Futures relating to development of its Outcomes Framework may provide more precise estimates to inform this.

Time periods

In terms of the number of periods with complete data, we assumed three pre-intervention and three post-intervention periods, corresponding to six years of data per college. This assumes that DFNPS has been rolled out for at least three years, which is feasible given that DFNPS has been operating since 2018 (DFN Charitable Foundation, 2019).

Treatment exposure

In terms of the proportion of treated colleges, DFNPS reports having worked with approximately 150 colleges since launch in 2018 (DFN Charitable Foundation, 2019), out of around 217 FE colleges in England at present (Higgingbotham, 2025), implying a treated share of 69% assuming a consistent total number of FE colleges. For the purposes of the power calculations, however, we adopted a more conservative assumption of a 50% treated share. This reflects the need for continuous and complete records across all years for each college, and staggered programme adoption across colleges.

Power calculation results

The results of the power calculations are summarised in Table 13. The values reported represent the number of colleges required in the matched sample for the DiD design to achieve 80% statistical power at the 5% significance level for each MDES scenario. All scenarios are based on six time periods, two-sided distribution, college-level analysis, untreated employment incidence at 5.9%, and 50% treated share.

Table 13: Power calculations for a college-level DiD

Parameters	Scenario 1	Scenario 2	Scenario 3
MDES (risk ratio, comparator rate: 5.9%)	2.7	6.08	11.2
MDE (percentage points change, comparator rate: 5.9%)	10	30	60
Total number of colleges required	121	13	4

While the scenarios presented in Table 13 vary only the MDES, changes in other parameters would also affect the required sample size. For example, if the proportion of treated colleges were lower than 50%, a larger number of colleges would be required to capture the same effect size, with the change being larger for the smallest MDE scenario (10pp). Similarly, if panel data with more than 6 time periods were available, the same MDES could be detected with fewer colleges, while keeping the same statistical power. Employment incidence also matters. For instance, if the baseline incidence employment rate were to go from 5.9% to 20%, the number of colleges required to detect a 30-percentage point effect would increase from 13 to 24.

Overall, the power calculations suggest it is feasible to construct a DiD design with sufficient power to detect effects above 10 percentage point difference on employment rate, given the assumed incidence rate and available population of

colleges. The estimated sample size requirements are considerably lower than both the total number of colleges in England (212) and the number of colleges that DFNPS reports having worked with (150).

However, several risks must be mitigated during implementation. First, it is key to avoid having only two time periods of data, as this would substantially increase the number of colleges required to detect a MDE of 10 percentage points.

Second, there is a risk of attenuation in the estimated effect size, arising both from averaging impacts over periods of stronger and weaker programme effects and from potential overlap between DFNPS provision and the changing BAU across time. If this attenuated effect is smaller than a 10 percentage points MDE, a larger number of colleges would be required to achieve statistical significance.

Finally, differences in registration methods across colleges and over time could reduce the MDES. Therefore, it is crucial to aim for the largest possible sample to mitigate potential noise and biases in the data. Where appropriate, the feasibility of imputation should be explored.

4.5.2. Acceptability and working relationships

DFNPS

Overall, DFNPS has shown strong interest in, and willingness to collaborate on, developing this evaluation design, particularly to provide historical information about their engagement with partner colleges.

There was some concern within the DFNPS team that the estimated effect might underestimate the impact of the programme because comparator colleges frequently run their own supported internship programmes. Consequently, any estimated impact would capture the effect of DFNPS relative to other supported internship programmes, rather than as a clear contrast with no provision. To address this, the feasibility of incorporating a control variable to account for the presence of other supported internships should be explored.

Again, the DiD method should be implemented alongside other evaluation approaches, and it will be important to continue discussions with DFNPS staff to clarify how each approach captures different dimensions of the effectiveness of supported internships, and how findings should be interpreted and communicated. In particular, this approach can explore how DFNPS compares to historical BAU practice, which in many settings may also involve other supported internships.

Notwithstanding these considerations, a college-level matched DiD approach is broadly acceptable to DFNPS.

Other stakeholders

Importantly, this design would not require establishing new relationships with comparator colleges, as the data required can be accessed directly by researchers through DfE administrative datasets. This points to one final consideration for the DiD – it is ultimately contingent on engagement from DfE to confirm and provide data access, and that it aligns with their priorities and any related evaluation activity already in progress.

4.5.3. Ethics

The evaluation would not affect any individuals' access to existing provision, therefore avoiding potential ethical concerns around this.

The proposed DiD approach would rely on existing administration datasets. As these datasets are already held within government systems in pseudonymised format and no new data would be collected for the purposes of the evaluation, the analysis does not raise significant additional ethical concerns. Moreover, the analysis would be conducted at the college-level, meaning that individual young people would not be directly identifiable or engaged as research participants.

4.5.4. Feasibility assessment summary

Table 14 below provides a summary of this method's feasibility. Overall, the college-level DiD design is methodologically feasible, provided that the parallel trends assumption holds and that barriers around data consistency and completeness can be worked around. It would not be possible to detect small effects, but moderate-to-large effect would be detectable given the extent of DFNPS' historical delivery. Conversations with DFNPS suggest that this method is broadly acceptable, acknowledging that its value as an evaluation on its' own is limited, and should be seen as a complement to other methodologies. Finally, minimal new working relationships are required, apart from DfE engagement in facilitating data access, since the analysis relies on administrative data. It is also ethically relatively straightforward, since only pseudonymised data will be accessed.

Table 14: Feasibility assessment summary

Approach	Robustness	Practical feasibility	Acceptability	Working relationships	Ethics
Matched Difference-in-					

Difference at college level					
-----------------------------	--	--	--	--	--

4.6. Regression Discontinuity Design

As part of this study, the feasibility of implementing both sharp and fuzzy RDD approaches was explored.

In the context of DFNPS, this study also explored different time horizons, including whether a robust RDD is feasible with the 2025 cohort – the cohort whose recruitment journey was captured by this feasibility study – as well as with forthcoming cohorts of participants (2026).

Overall, there are notable barriers and risks of this methodology, as presented in this section, which limits its feasibility and robustness. Ultimately, this section will recommend against using an RDD design for future evaluation, but – unlike other evaluation designs that were ruled out at an earlier stage and described in the appendix – an RDD design was explored in more depth due to initially showing more promise, and therefore is included in the main body of the report.

4.6.1. Practical feasibility and robustness

This section explores the RDD approach in more depth, including more information and considerations about the methodology, enablers, barriers, and required sample sizes. Overall, it finds that there are notable limitations related to feasibility and robustness.

Methodology

A sharp RDD would operate as follows. Young people attend site selection days, during which information is collected to construct the running variable – specifically, the total score derived from the rubric sheet. A pre-agreed and arbitrary cut-off rubric score then determines who is offered a supported internship and who is not.

A pre-determined bandwidth is defined around this cut-off to identify the individuals used within the analysis. The bandwidth must be wide enough to include enough individuals to provide adequate statistical power, yet narrow enough to ensure that individuals just above and just below the cut-off are highly comparable, such that the allocation of supported internships can be considered “as good as random.”

Identification assumptions and bandwidth selection

For an RDD to be valid, the running variable must be continuously distributed around the threshold. Under this condition, any jump in outcomes at the cut-off can be causally attributed to the supported internship.

Keeping the bandwidth relatively tight helps ensure that individuals just above and below the cut-off are sufficiently comparable, such that differences in outcomes can be more confidently attributed to the supported internship rather than differences in underlying characteristics. However, there is a trade-off between selecting a narrow bandwidth, which improves the validity of the design, and selecting a wider bandwidth, which increases the number of observations available for analysis and therefore statistical power. Identifying the optimal bandwidth is typically an iterative process involving statistical tests and robustness checks, with the final choice depending on the characteristics of the data.

As a result, not all individuals who attend a site selection day – or who are eligible for the programme more broadly – would necessarily be included in the final analytic sample for a full-scale RDD, as only those within the selected bandwidth contribute to the analysis. The size of the optimal bandwidth can vary substantially depending on the distribution of the running variable and the available sample size. For example, studies in educational settings have found that optimal bandwidth selection may exclude a substantial proportion of observations (Deke & Dragoset, 2012), with standard practice often involving testing multiple bandwidth specifications and retaining fewer than 50% of the original sample. This highlights the importance of ensuring that there are sufficient numbers of participants close to the cut-off to maintain adequate statistical power.

Pooling cohorts

The feasibility study considered whether a robust RDD could be implemented using only the 2025 cohort, or whether pooling data from future cohorts, particularly 2026 cohort, would strengthen the design.

Pooling multiple DFNPS cohorts offers a clear potential advantage: increased sample sizes would improve statistical power and enhance the robustness of the RDD estimates. In principle, this could lead to stronger and more reliable evidence on the programme's impact.

However, the feasibility of this approach depends on several factors, including Youth Futures preferred timelines for publishing findings and the availability of funding. Pooling cohorts also has important implications for outcome measurement and evaluation duration. Sufficient time must be allowed both for participants to engage in job-search activities and for meaningful outcome data to be collected following the completion of

internships. While extending the sample size is convenient from a statistical power perspective, it is important to mention that in the context of this study, pooling cohorts beyond 2026 would be complicated as their outputs would be available beyond the window of the intended evaluation.

Enablers

Availability of a running variable

A key enabler for conducting an RDD of DFNPS is the presence of the rubric score embedded within the selection process, which can serve as the running variable. This score is a structured index applied by assessors across sites and site selection days to inform treatment allocation.

During the feasibility study, King's were able to work closely with DFNPS to access this data. The analysis of the 2025 cohort shows that the rubric data is highly complete, and DFNPS has been proactive in addressing data quality queries as they have arisen.

Existence of a cut-off

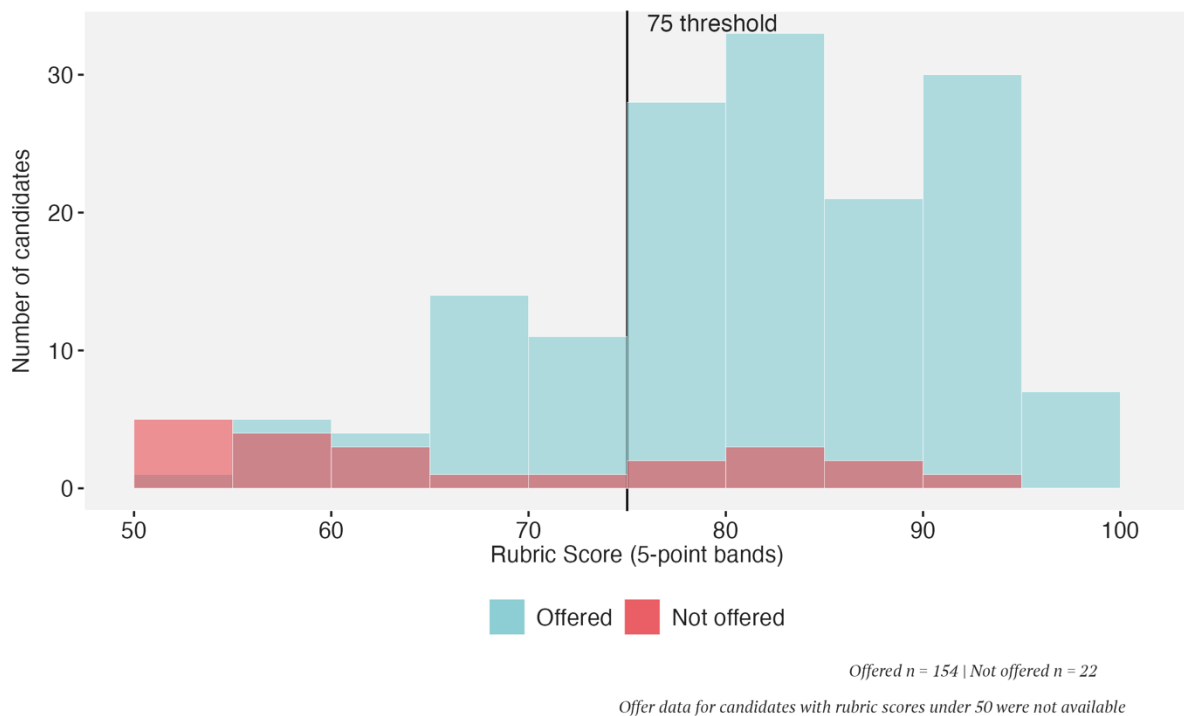
Another important condition for an RDD is the possibility of defining a cut-off rubric score that determines whether a young person is offered an internship offer. This threshold could either be newly defined for evaluation purpose or based on existing, allocation rules.

To assess whether such a natural threshold exists in practice, the feasibility study examined data from the 2025 cohort. Of the full sample of 399 young people captured in the rubric data, data was used for the 179 young people where there was data available for both rubric-survey scores and internship offers. Among these, 157 (88%) were offered a supported internship, with rubric scores ranging from 55 to 100. Multiple Imputation by Chained Equations (Azur et al., 2011; Wilson, 2021) was used to predict missing supported internship offer outcomes using the information in the rubric sheets and baseline survey responses (including age, gender, previous employment experiences, and outcomes at baseline). It is important to stress that this imputation was carried out solely to generate indicative figures for the feasibility study and should be interpreted with extreme caution, as more than half of the sample is imputed, substantially limiting the reliability of any conclusions.

Using the imputed sample of 377⁹ young people, it is predicted that 294 (78%) were offered a supported internship. Analysis of the imputed rubric score distribution reveals a discontinuity at a score of 75, as shown in Figure 7. The 62 candidates scoring between 37 and 74 had a 55% chance of receiving an offer, compared with 88% among the 231 candidates scoring between 75 and 100.

While this represents a jump in the probability of being offered an internship, offers and non-offers are observed both above and below the threshold, indicating there is no sharp threshold. This pattern suggests that a fuzzy RDD would be a more appropriate design.

Figure 7: Rubric score distribution for all young people score around threshold of 75



The presence of a natural discontinuity at 75 has both positive and negative implications for the robustness and feasibility of an RDD design. On the positive side, using an existing allocation threshold reduces the risk of resistance from site assessors as it aligns broadly

⁹ Approximately 52% of missing data had to be imputed, so results should be interpreted with caution, still within margins of what can be cautiously still considered sufficiently accurate and reliable – see Junaid et al. (2025).

with observed practice rather than imposing an artificial rule. However, there are also negative implications, described below in the section on barriers.

Barriers

Violation of the continuity assumption in the running variable

While the presence of a natural discontinuity in the distribution of rubric scores helps indicate where a threshold could be set with minimal disruption, it may also point to a violation of a critical identification assumption of an RDD. For an arbitrary allocation threshold to effectively divide the relevant sample (within a defined bandwidth) into comparable groups, the running variable (the rubric score) should follow a smooth and continuous distribution. Ideally, the rubric score identified in this feasibility study would present as smooth and continuous, allowing an arbitrary threshold to be set without introducing discontinuities that could bias any subsequent estimation.

Given that a discontinuity does exist in the rubric score distribution, it was important to test whether its extent, meaning how pronounced it is, poses a genuine problem for an RDD design. To formally test this, we conducted a McCrary test (Cattaneo, Jansson & Ma, 2020) that assumed an arbitrary threshold where the rubric score equals 75. The test rejected the null hypothesis that the density of the distribution is continuous at the threshold, using the Mean Squared Error (MSE) optimal bandwidth, which is selected to balance estimation bias and variance at the cutoff. As a sensitivity check, we re-ran the test with a restricted bandwidth of ± 10 ; this specification was non-significant ($p = 0.386$), though we consider this likely reflects reduced statistical power at the narrower bandwidth, rather than evidence that no discontinuity exists in the underlying density. The results of both tests are presented in Appendix I.

There are some possible explanations for this discontinuity. First, as McCrary (2008) points out, such breaks in the density around the proposed cut-off are typically interpreted as a sign of sorting or manipulation. In this context, this would suggest that rubric scores tend to be inflated, even unintentionally, to ensure young people qualify for the internship programme. A second, equally plausible explanation is that the true underlying density function of the rubric score is not linear and instead follows a curvature that is being masked as a discontinuity detected by the test given the current sample size, particularly since RDDs are designs that benefit from larger sample sizes.

If a full-scale RDD were to be conducted, it would be important to establish which explanation prevails. While internal college screening prior to assessment days likely skews the overall rubric score distribution towards higher-suitability candidates, it does not plausibly explain the discontinuity detected. The qualitative evidence, meanwhile, does not explicitly confirm deliberate on-the-day manipulation of scores. It did note,

however, that rubric scores may mask some inconsistencies in marking practices, as flagged by site assessors during interviews, which could be contributing to the discontinuity observed.

Regardless of the explanation, if individuals just above and just below the threshold are not locally comparable, any observed discontinuity in outcomes cannot be causally attributed to the programme. As a result, a robust RDD using solely the 2025 cohort data alone is not feasible, as it is not possible to rule out the risk of sorting nor is the sample large enough to unmask alternative distributional shapes.

Regarding an RDD with multiple pooled cohorts, its feasibility would depend on the plausibility of ruling out the risk of sorting. Given the issues with the 2025 cohort, any attempt to use this data would require further qualitative work specifically focused on detecting strategic/sorting behaviour to rule it out. For subsequent waves, it would be important to implement necessary changes to the assessment process to safeguard it against manipulation, ensuring any changes remain consistent across waves to ensure comparability.

Limited comparability around the cut-off

Another important condition for a robust RDD is that individuals above and below the cut-off are comparable in observable characteristics within an optimally defined bandwidth. In a full-scale evaluation, this would involve defining a narrow range around the threshold where treatment assignment is considered "as good as random," thereby ensuring balance between the groups.

Ideally, King's would formally test for balance across key characteristics within this optimal and narrow bandwidth. However, given current sample size constraints, restricting the analysis to a narrow bandwidth would further diminish the analytical sample. Therefore, for the purpose of this feasibility report, all observations above and below the cutoff were compared to assess broad pre-treatment differences between the two groups.

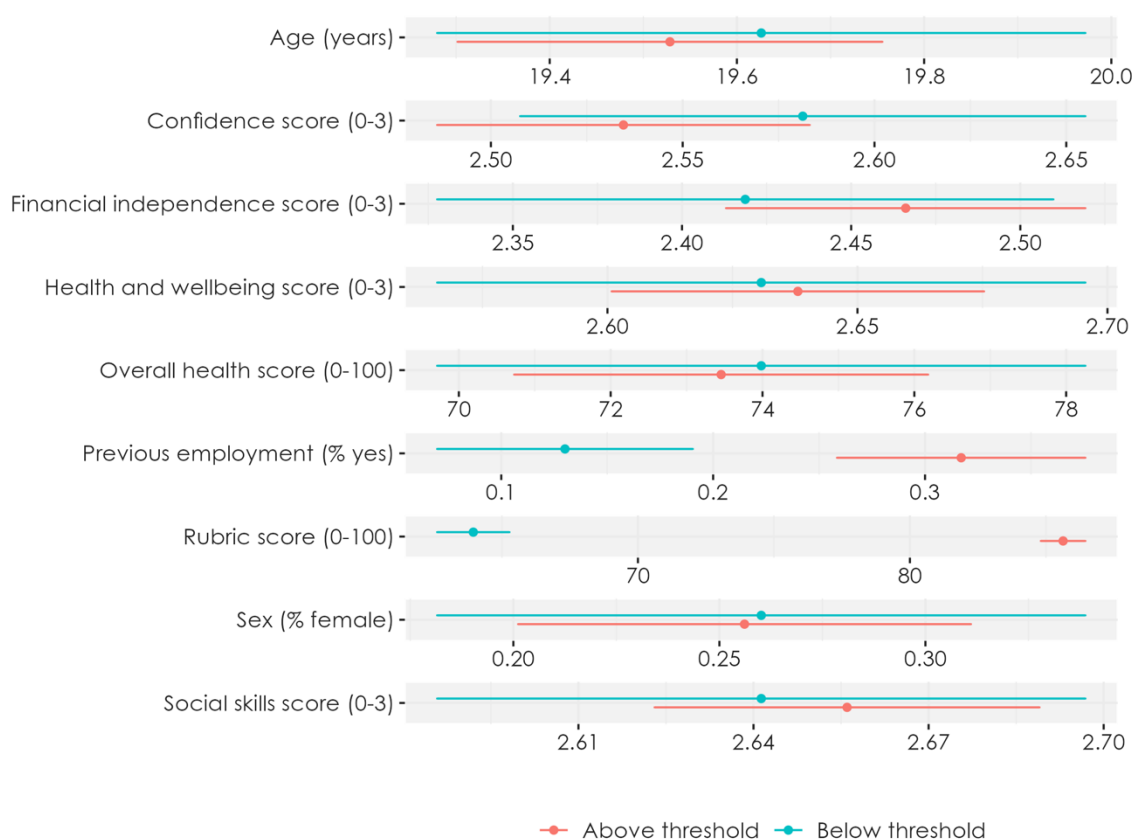
Indicative balance checks¹⁰ **Error! Reference source not found.** show that candidates above the threshold were significantly more likely to report prior employment experience in the baseline survey. All other observable characteristics examined did not differ

¹⁰ While ideally balance checks were to be tested within bandwidths, given sample size limitations that may undermine power, we kept these as indicative on the overall sample. In this sense, imbalances detected are likely more pronounced if the sample size were to be closer to a threshold.

significantly between groups. While this suggests that the groups are not radically different, the detected difference reinforces the need for a narrow bandwidth in any future RDD.

It is also important to note that previous employment experience is highly linked with the rubric sheet score itself – the running variable – as this captures work readiness. As such, this difference may be less concerning than differences in unrelated characteristics. Nevertheless, a difference detected in observable characteristics across groups suggest there may also be potential differences in unobservable characteristics, so finding an optimal bandwidth would not be trivial. It's also possible that in some cases, there was not enough power to detect differences on variables that are currently showed as balanced.

Figure 8: Balance checks of participants above and below the 75 threshold



Limited size of comparison group

A key concern identified during the feasibility study is the low number of candidates attending site assessment days who are not offered a supported internship, that could

then be used to construct the comparison group. Observations and conversations with site assessors indicate that most candidates who attend these days are well-prepared and subsequently accepted into the programme. Consequently, there is a limited number of candidates not assigned to an internship to construct the comparison group from, restricting the statistical power of the analysis. Based on the 2025 data, only 22 individuals (12%) out of 179 (for which we had internship offer data) were not offered an internship. In the imputed dataset, this rises to 83 (22%), which is still relatively small. Under a fuzzy RDD with the threshold of 75, only 113 candidates (30%) would fall under the cut-off.

By comparison, RDD evaluations of employment-related programmes often report around 50% of observations under the threshold within a pre-defined bandwidth (Kaaksonen et al., 2022; Charaja, 2023). For instance, Charaja (2023) used a sample size of 8,158 to establish two bandwidth scenarios of 850 and 499 observations respectively, with 52% of observations allocated under the cut off. However, a study in a different field on information campaigns reported a share of 27% under the threshold within pre-defined bandwidth, but with a much larger total sample of 7,760 observations within the bandwidth, 2,104 of which (27%) were under the cut-off and hence not treated following a sharp RDD (Tyllianakis et al., 2025). Overall, the limited size of the potential comparison group in DFNPS substantially restricts statistical power.

Insufficient density

A further barrier relates to the density of observations around the cut-off. A robust RDD requires a sufficiently dense distribution of data points closely around the cut-off to validate the linearity of the running variable around the threshold and to test several possible bandwidths. The current 2025 cohort sample is not dense enough around the threshold to reliably validate the linearity assumption and test bandwidths. This would reduce the likelihood of generating a continuous running variable suitable for RDD analysis.

Limited policy relevance of LATE estimates

Finally, RDDs estimate the impact of the intervention only for individuals close to the cut-off, an estimation known as the Local Average Treatment Effect (LATE). This may not reflect the impact of the programme on the full participant population, which would likely be of greater policy relevance to DFNPS, Youth Futures, and other stakeholders such as DfE.

While this limitation is inherent to the RDD framework – and is necessary to ensure comparability – it represents a substantive trade-off in this context. DFNPS has expressed interest in understanding the broader impact of the programme (i.e., through estimation

of the Average Treatment Effect or ATE). An RDD requiring a very narrow bandwidth (and thereby population for which the impact of DFNPS is estimated) may therefore be misaligned with the broader objectives of the evaluation.

Sample size

Power calculations were conducted to assess the MDES that could be identified using the current 2025 cohort. These calculations are intended to evaluate whether an RDD would be sufficiently powered to detect meaningful programme effects under different bandwidths and sample size scenarios.

The calculations are based on the imputed dataset of 377 candidates for whom rubric scores and survey data were available. Power calculations for RDDs differ from those used in other evaluation designs. They are data-driven and rely on simulations conducted over a seed dataset – in this case, the imputed 2025 dataset. Simulations were performed using the 'rdpower' package in R.

The key parameters specified in the simulations were the MDES, the bandwidth around the cut-off point, and the cut-off point itself.

The parameter of interest in these simulations is statistical power. A statistical power of 80% is commonly accepted for an evaluation design to be considered reliable, where designs with a lower statistical power are unlikely to produce robust results.

Given the absence of data on employment outcomes, internship offers are used as a proxy outcome. Accordingly, the MDE reported in the power calculations are interpreted as percentage-point changes in the probability of employment.

Four illustrative scenarios were examined:

- Scenario 1: a +/- 25-point bandwidth around the threshold i.e., covering all individuals above the threshold.
- Scenario 2: a +/- 38-point bandwidth around the threshold i.e., covering all individuals below the threshold;
- Scenario 3: a +/- 5-point bandwidth around the threshold, demonstrating the limitations of a narrow bandwidths using the current cohort.
- Scenario 4: a simulated pooled sample with twice the size of the 2025 cohort, using a +/- 5-point bandwidth.

The MDE used across all estimations is 56 percentage points, corresponding to the estimated effect from the Project SEARCH programme evaluated in the US. The cut-off point is set at 75, consistent with observed allocation patterns in the 2025 cohort.

The results are summarised in Table 15.

Table 15: Power calculations for a Regression Discontinuity Design

Parameters	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Bandwidth (in rubric score points)	+/-25	+/-38	+/-5	+/-5
Total sample size	377	377	377	754
Sample size within the bandwidth	374 (30% under cut-off)	377 (31% under cut-off)	100 (21% under cut-off)	200 (21% under cut-off)
MDE (change in likelihood of employment measured in pp)	56 pp	56pp	56pp	56 pp
Statistical power	0.88	0.93	0.11	0.17

Scenario 1 and 2 achieve statistical power above the 80% threshold, with power estimates of 0.88 and 0.93, respectively. However, these scenarios rely on very wide bandwidths, encompassing most of the sample.

Scenarios 3 and 4 indicate that when narrower bandwidths are used, those more likely to generate balanced groups, statistical power drops sharply. Power falls to 0.11 under Scenario 3 and to 0.17 under Scenario 4, even when the sample is doubled. These results indicate that even doubling up the sample achieved for 2025 cohort is insufficient to overcome the limitations imposed by low density around the cut-off.

Potential strategies to improve power

Several factors could, in principle, facilitate the detection of smaller effect sizes and allow for narrower bandwidths.

First, the use of administrative datasets capturing employment outcomes could allow inclusion of all 438 candidates for whom some form of data exist in the 2025 cohort, and potentially additional cohorts.

Second, a set of covariates could be included from administrative data to improve the precision of the estimator by controlling for observable differences across groups.

However, even under these more favourable conditions, gains in statistical power are expected to be limited. This is illustrated in Scenario 4, where doubling the sample size result in only modest improvements in power.

Overall, the power calculations indicate that bandwidths narrower than ± 25 percentage points around the cut-off (i.e., within the region of being sufficiently narrow to make those above and below the threshold sufficiently comparable) produce designs with statistical power well below the conventional 80% threshold, rendering them unreliable for detecting meaningful effects. This finding holds both for analysis based on the 2025 cohort alone and for simulated pooled samples.

In addition, we also run the Imbens–Kalyanaraman (IK) bandwidth selection procedure on a dataset with available rubric scores and internship offers (imputed where information on offers was not available). The procedure returned an optimal bandwidth of 36 points either side of the suggested rubric cut-off (75), which spans nearly the full observed range of rubric scores and exceeds its upper bound. A bandwidth this wide is diagnostic of insufficient observation density around the cut-off: the MSE-optimal procedure is forced to widen the window aggressively to control estimation variance, in turn undermining the local exchangeability assumption on which RDD identification depends. Under current conditions, this suggests it would not be possible to construct a sufficiently local sample to support robust RDD estimation. The full IK output is reported in Appendix I.

4.6.2. Acceptability

DFNPS staff have expressed general support for this approach, especially noting that RDDs would avoid some of the potential ethical concerns associated with randomisation. In contrast to randomisation, an RDD aligns more closely with existing selection practices, which enhances its perceived ethical acceptability.

However, DFNPS already raised substantive concerns about the proposed design. A key concern relates to the size of the comparator group. Given that candidates who attend site assessment days are typically ready to be in a supported internship, the number of people falling below any allocation threshold is likely to be low. This raises doubts about whether a sufficiently large and credible comparator group could be constructed.

Another concern raised is that an RDD would estimate programme impacts only for participants located close to the cut-off rubric score (i.e., the LATE). As such, the impact estimates may not adequately reflect the impact of DFNPS on the wider group they serve. DFNPS indicated a preference for impact estimates that capture outcomes for all eligible participants, rather than a narrow subgroup.

These two concerns relate to structural features of an RDD methodology and are therefore not straightforward to address. As a result, they limit the overall acceptability of an RDD.

There may be some scope to mitigate these issues in a full-scale evaluation. Potential options include: pooling multiple cohorts and potentially extending recruitment, modifying the referral process so that any screening conducted by colleges is light-touch, or by establishing an arbitrary cut-off that yields a more appropriate sample composition. While DFNPS stakeholders were not necessarily strongly opposed to these approaches, it is unlikely that this methodology would be accepted as a standalone design and would instead be more feasible as a complement to other approaches, such as an RCT.

4.6.3. Working relationships

DFNPS

An established and constructive working relationship with DFNPS represents an important enabler for a potential full-scale RDD. Core to this design is the availability of consistent and high-quality rubric score data. As these scores are collected by site assessors during selection days, DFNPS is a key partner in ensuring that data is accessible and collected consistently across sites.

During the feasibility study, DFNPS demonstrated a high level of engagement by proactively supporting efforts to recover missing rubric scores. In addition, DFNPS staff were highly responsive in providing additional information to facilitate matching of programme data to administrative records on employment outcomes. This level of collaboration is a clear strength and would be particularly valuable if a pooled analysis across the 2025 cohort and future cohorts were pursued.

Colleges

Relationships with colleges are also key in referring young people to site assessment days, and therefore are influential in shaping the size and composition of the sample. For a full-scale evaluation, further work would be required to encourage colleges to increase the number of students referred to site selection days, and to ensure that employment outcome data are captured for young people who are assessed but are not offered DFNPS. Conversations with colleges during the feasibility study indicate that such engagement is feasible, but would require additional coordination and collaboration with colleges.

4.6.4. Ethics

Participation in the evaluation would not affect any individuals' access to existing services or supported internships, avoiding ethical issues around withholding access to programme participation. All allocation decisions would continue to be made through standard DFNPS processes, independent of evaluation participation.

Nevertheless, standard ethical requirements would still apply to the analysis and reporting of the existing 2025 cohort data, and on the potential collection, analysis, and reporting from the 2026 cohort data as well. Similar to the 2025 cohort data collection, the 2026 cohort study would need to gain informed consent from participants, including consent for administrative data linkage.

A specific ethical consideration relevant to an RDD design is that the primary impact analysis focuses only on individuals whose rubric scores fall within a specific bandwidth around the cut-off point. As a result, some young people from whom data have already been collected in the feasibility study – or would be collected in future cohorts – may ultimately not be included in the final impact estimates because their rubric scores lie outside this range.

While this exclusion is methodologically necessary for the RDD, it raises a relatively minor ethical consideration related to participant burden and expectations. In particular, some participants may have invested time and effort in providing data without directly contributing to the primary impact analysis.

To mitigate this, all data would still be used appropriately; for example, data outside the bandwidth would still contribute to descriptive analysis and to the IPE, thereby improving contextual understanding of programme delivery and participant experiences.

4.6.5. Feasibility assessment summary

Table 16 summarises the feasibility assessment of two different RDDs, the first using only the data from the cohort 2025 (for which rubric scores are already held as part of the feasibility study), and the second pooling data across the 2025 and 2026 cohorts.

An RDD's feasibility is constrained by several interrelated factors. These include the small number of candidates who are not offered an internship, evidence of likely sorting around the threshold, and an overall sample size that does not allow for defining a sufficiently narrow bandwidth. For an RDD based solely on the 2025 cohort, these risks are particularly high. The available sample is not large or dense enough around the threshold to identify a bandwidth that would plausibly produce balanced groups below and above the threshold. As a result, any estimated effects would be highly vulnerable

to bias. Even as a pilot, an RDD using only the 2025 cohort is unlikely to generate insights of sufficient robustness or policy relevance to meaningfully contribute to the existing evidence base.

Expanding the design to include additional cohorts could address some of these limitations, particularly if steps are taken to define the internship allocation threshold and manage the risk of sorting. However, concerns about data density remain. If the distribution of rubric scores observed in 2025 persists in future cohorts, achieving the sample size required for an optimal bandwidth and well-balanced groups would continue to be challenging. Addressing this would likely require efforts to increase the density of observations around the threshold, which could imply altering the DFNPS delivery or selection model. Such changes, in turn, would raise questions about the comparability of the different cohorts pooled for analysis.

From an ethical and operational standpoint, DFNPS considers this design to be acceptable. The strong existing working relationships are a key enabler for implementing the methodology, though uncertainty remains regarding the capacity of colleges to refer a suitable number of candidates to assessment days.

Table 16: Feasibility assessment summary

Approach	Robustness	Practical feasibility	Acceptability	Working relationships	Ethics
Regression Discontinuity Design (2025 cohort)					
Regression Discontinuity Design (pooled cohorts)					

5. Conclusions

This report described the findings of a feasibility study exploring the suitability of DFNPS for a future impact evaluation, including an assessment of the feasibility of various specific evaluation designs. This concluding section summarises the key findings on programme delivery that factor into the feasibility of potential impact evaluation designs, and then outlines the recommended approaches for evaluating DFNPS's impact on participants' employment outcomes.

5.1. Programme recruitment and set up

Observations of site assessment days, where candidates' suitability for a supported internship is assessed, revealed significant variation in format, delivery, and available facilities across sites. Assessment days were typically held in colleges or Local Authority buildings and structured around carousel-style task rotations and one-to-one interviews. Tasks were designed to reflect the practical skills required for prospective supported internship roles, and generally fell within candidates' capability ranges, with assessors providing appropriate support when needed. A total of 68 site assessment days were completed during the timeframe of this study, and a total of 406 candidates completed the research survey.

Site assessors interviewed reported that, at times, inadequate space at some sites created distractions and limited privacy, occasionally reducing candidates' comfort and task performance. Conversely, according to site assessors and observations completed by the King's evaluation team, sites with clear structures and adequate space supported a smoother process and more consistent candidate engagement and assessment.

The rubric sheet – the main tool for scoring and assessing candidates' suitability – was widely used but inconsistently applied across assessors and sites. Misunderstandings of scoring criteria led to variability in candidate assessments, and manual entry during roundtable discussions slowed decision-making. Nonetheless, the rubric provided a valuable framework for ensuring structure and transparency, especially when communicating selection decisions to families.

Candidate selection decisions were made collectively during roundtable discussions, where assessors reviewed task performance, rubric sheet scores, and their broader qualitative assessment of candidates. While discussions were generally constructive and collaborative, time pressures sometimes affected information recall and consistency in

candidates' assessments. Host businesses valued this participatory approach, noting that it increased ownership and alignment between colleges and employers.

Overall, while the research identifies potential to improve consistency across site assessment days, the current candidate selection process – including the site assessment days, scoring through rubric sheets, and assessors' discussions – appears broadly robust and facilitates the implementation of an impact evaluation.

5.2. Feasibility assessment findings

The study assessed the feasibility of several different impact evaluation designs to estimate the effect of DFNPS on young people's employment outcomes, including RCTs, DiDs, and RDDs. The appendix also discusses three additional approaches that were ruled out at an earlier stage of the feasibility study. Our overall assessment of the different methodological options is summarised below.

Based on the evidence gathered and discussions with DFNPS and Youth Futures, an **RCT with DFNPS scaled up to SEND learners without EHCPs** would be the most robust and feasible design to causally evaluate the impact of DFNPS on young people's employment outcomes. This approach offers strong internal validity and is well aligned with DFNPS's existing operational model and referral processes. The design is strengthened by the availability of a large pool of SEND learners without EHCPs that should be broadly comparable to DFNPS' usual target cohort, enabling the recruitment of a sufficiently large sample for the trial, which should then produce highly generalisable findings. Importantly, this design is ethically sound, as it expands access to supported internships through scaled-up delivery rather than withholding opportunities from the current target cohort. The main practical enablers are the strong support of DFNPS' leadership and established partnerships with colleges and host businesses, alongside their ability to leverage these relationships to secure new placements for trial participants.

A **college-level matched DiD design** is also a feasible and acceptable approach to evaluate DFNPS. Methodologically, its robustness depends on meeting the parallel trends assumption and on the possibility of accessing standardised (or standardisable) administrative data. While this approach would not allow for the detection of small effects, it could identify moderate to large effects (which are within a plausible region for DFNPS based on results from the US trial), and the approach poses no ethical concerns as it relies solely on anonymised administrative data. However, this option is quite a blunt instrument analytically as it fails to capture the individual participant context, the rubric score, and work-readiness. Instead, it simply assesses the extent to which a college implementing DFNPS sees a change in the aggregate employment

outcomes of eligible students compared to the comparator group. We would therefore recommend it would be best deployed in combination with another approach (e.g., an RCT) alongside a strong IPE.

An **RCT with scale-up to new colleges** also represents a robust and technically feasible design. It would allow for causal attribution through individual-level randomisation within new settings, without disrupting DFNPS' existing delivery. However, this approach presents greater operational and resource challenges for DFNPS, particularly around recruiting new colleges and employers, securing data-sharing agreements, and achieving the large sample size required to detect moderate effects, and especially given the need to start a potential RCT with the 2026 cohort due to Youth Futures funding timelines. While feasible in principle, conversations with DFNPS indicate some reservations about its practical deliverability and the associated cost implications. Nonetheless, this approach remains ethically sound and consistent with DFNPS' mission.

Finally, a **RDD approach** would face substantial methodological and practical barriers. These include a small number of candidates who are not offered an internship which limits the size of the potential comparison group, evidence of likely sorting around the threshold, and an overall sample size that does not allow for defining a sufficiently narrow bandwidth around the threshold and which risks producing unbalanced groups. As a result, any estimated effects would be highly vulnerable to bias, and unlikely to be robust enough to provide meaningful insights. Although other criteria were met, such as being ethically acceptable and having strong working relationships with stakeholders, the feasibility study ultimately concludes that an RDD is not a suitable approach for a future impact evaluation of DFNPS.

6. References

- Adams, L., Tindle, A., Basran, S., Dobie, S., Thomson, D., Robinson, D., & Shepherd, C. (2017). Experiences of education, health and care plans: A survey of parents and young people.
- Adu, A., & correspondent, A. A. P. (2025). Children with special needs in England may lose legal right to school support plans. *The Guardian*.
- Ashworth, M., Heasman, B., Crane, L., & Remington, A. (2024). Evaluating a new supported employment internship programme for autistic young adults without intellectual disability. *Autism*, 28(8), 1934-1946.
- Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res*. 2011 Mar;20(1):40-9. doi: 10.1002/mpr.329. PMID: 21499542; PMCID: PMC3074241.
- Barham, C., Walling, A., Clancy, G., Hicks, S., & Conn, S. (2009). Young people and the labour market. *Economic & Labour Market Review*, 3(4), 17-29.
- Camden, B. (2024, February 23). *Destinations unknown: ILR data cut angers MCAs*. FE Week. <https://feweek.co.uk/destinations-unknown-ilr-data-cut-angers-mcas/>
- Cattaneo, M. D., Jansson, M., & Ma, X. (2020). Simple local polynomial density estimators. *Journal of the American Statistical Association*, 115(531), 1449–1455. <https://doi.org/10.1080/01621459.2019.1635480>
- Charaja, L. I. (2023). The employment effect of subsidized transit on unemployment assistance recipients. <https://ledainga.github.io/uploads/madrid.pdf>
- Christensen, J. J. (2015). Longitudinal outcomes of Project SEARCH in upstate New York. *Journal of Vocational Rehabilitation*, 42(3), 247–255.
- Deke, John & Dragoset, Lisa. (2012). Statistical Power for Regression Discontinuity Designs in Education: Empirical Estimates of Design Effects Relative to Randomized Controlled Trials. Princeton, NJ: *Mathematica Policy Research*. ERIC Document No. ED533141
- DFN Charitable Foundation. (2019). Written evidence submitted by the DFN Charitable Foundation (SCN0714). UK Parliament, House of Commons Education Committee.
- Department for Education (2013). Accessing LEO data via the ONS SRS Frequently Asked Questions. [online] GOV.UK. Available at:

[https://assets.publishing.service.gov.uk/media/65426ab61f1a600010360bfc/Accessing_L
EO via ONS SRS I2 FAQ ver 1.3.pdf#:~:text=pseudonymised%20organisation%2Dlevel%
20data%20can%20now%20be%20linked,be%20pseudonymised%20before%20it%20can%
20be%20linked](https://assets.publishing.service.gov.uk/media/65426ab61f1a600010360bfc/Accessing_LEO_via_ONS_SRS_I2_FAQ_ver_1.3.pdf#:~:text=pseudonymised%20organisation%2Dlevel%20data%20can%20now%20be%20linked,be%20pseudonymised%20before%20it%20can%20be%20linked).

Department for Education (2024). Evaluation of the strengthening of the supported internship programme. GOV.UK. Available at:
[https://assets.publishing.service.gov.uk/media/672a1700fbd69e1861921b68/Supported_i
nternship interim report.pdf](https://assets.publishing.service.gov.uk/media/672a1700fbd69e1861921b68/Supported_internship_interim_report.pdf)

Department for Education (2015). *Special Educational Needs and Disability Code of Practice: 0 to 25 Years*. [online] GOV.UK. Available at:
[https://assets.publishing.service.gov.uk/media/5a7dcb85ed915d2ac884d995/SEND_Cod
e of Practice January 2015.pdf](https://assets.publishing.service.gov.uk/media/5a7dcb85ed915d2ac884d995/SEND_Code_of_Practice_January_2015.pdf)

Department for Education (2025). Supported internships. GOV.UK. Retrieved from:
[https://www.gov.uk/government/publications/supported-internships-for-young-people-
with-learning-difficulties/supported-internships](https://www.gov.uk/government/publications/supported-internships-for-young-people-with-learning-difficulties/supported-internships)

Department for Education. (2025a). Education, health and care plans, Reporting year 2025. GOV.UK. Retrieved from [https://explore-education-statistics.service.gov.uk/find-
statistics/education-health-and-care-plans/2025](https://explore-education-statistics.service.gov.uk/find-statistics/education-health-and-care-plans/2025)

Department for Education. (2025b). Special educational needs in England: Academic year 2024/25. GOV.UK. [https://explore-education-statistics.service.gov.uk/find-
statistics/special-educational-needs-in-england/2024-25](https://explore-education-statistics.service.gov.uk/find-statistics/special-educational-needs-in-england/2024-25)

Department for Education. (2025c). *NCCIS management information requirement: 2026 to 2027*. GOV.UK.
[https://assets.publishing.service.gov.uk/media/6943ca108f4636fa2c547e25/NCCIS_man
agement information requirement 2026 to 2027.pdf](https://assets.publishing.service.gov.uk/media/6943ca108f4636fa2c547e25/NCCIS_management_information_requirement_2026_to_2027.pdf)

Department for Education. (2026a, June 25). *Education, health and care plans: Reporting year 2026*. Explore Education Statistics. [https://explore-education-
statistics.service.gov.uk/find-statistics/education-health-and-care-plans/2026](https://explore-education-statistics.service.gov.uk/find-statistics/education-health-and-care-plans/2026)

Department for Education. (2026b, June 11). *Special educational needs in England: Academic year 2025/26*. Explore Education Statistics. [https://explore-education-
statistics.service.gov.uk/find-statistics/special-educational-needs-in-england/2025-26](https://explore-education-statistics.service.gov.uk/find-statistics/special-educational-needs-in-england/2025-26)

Folguera Bellmunt, C. (2025). Supported internships: vocational training and labor inclusion for students with intellectual and developmental disabilities. Equality, Diversity

and Inclusion: An International Journal. <https://doi.org/10.1108/EDI-05-2024-0218EtiSiLih>. DOI: <https://doi.org/10.1108/EDI-05-2024-0218>.

Gale, N. K., Heath, G., Cameron, E., Rashid, S., & Redwood, S. (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Medical Research Methodology*, 13, Article 117. <https://doi.org/10.1186/1471-2288-13-117>

Gedikli, C., Miraglia, M., Connolly, S., Bryan, M., & Watson, D. (2023). The relationship between unemployment and wellbeing: an updated meta-analysis of longitudinal evidence. *European Journal of Work and Organizational Psychology*, 32(1), 128-144.

GOV.UK (2025). *High needs funding: 2025 to 2026 operational guide*. [online] GOV.UK. Available at: <https://www.gov.uk/government/publications/high-needs-funding-arrangements-2025-to-2026/high-needs-funding-2025-to-2026-operational-guide>.

Hanson, J., Robinson, D., & Codina, G. (2021). Supported internships as a vehicle for social inclusion. *British journal of learning disabilities*, 49(4), 433-444.

Higginbotham, D. (2025, July). Overview of the UK's further education sector. Prospects. Retrieved from <https://www.prospects.ac.uk/further-education/overview-of-the-uks-further-education-sector/>

Inside Government. (2025). *Employment for young people with SEND*. Inside Government. <https://blog.insidegovernment.co.uk/schools/employment-for-young-people-with-send>

Institute for Employment Studies & Youth Futures Foundation. (2025). *Education to employment transitions support in 2024*. Youth Futures Foundation. <https://youthfuturesfoundation.org/wp-content/uploads/2026/02/Research-report-Education-to-employment-transitions-support-in-2024-December-2025-IES-Youth-Futures.pdf>

Jacob, U. S., Pillay, J., Adeoye, O. E., & Oni, T. K. (2023). Barriers to employment of people with intellectual disability. *Work*, 74(1), 207-218.

Junaid, K. P., Kiran, T., Gupta, M., Kishore, K., & Siwatch, S. (2025). How much missing data is too much to impute for longitudinal health indicators? A preliminary guideline for the choice of the extent of missing proportion to impute with multiple imputation by chained equations. *Population health metrics*, 23(1), 2.

Jones, S. C., Gordon, C. S., Akram, M., Murphy, N., & Sharkie, F. (2022). Inclusion, exclusion and isolation of autistic people: Community attitudes and autistic people's experiences. *Journal of Autism and Developmental Disorders*, 52(3), 1131-1142.

Laaksonen, M., Ilmakunnas, I., & Tuominen, S. (2022). The impact of vocational rehabilitation on employment outcomes: A regression discontinuity approach. *Scandinavian Journal of Work, Environment & Health*, 48(6), 498.

Leda Inga Charaja. (2023). The employment effect of subsidized transit on unemployment assistance recipients. Available at:
<https://ledainga.github.io/uploads/madrid.pdf>

Long, R. & Roberts, N. (2025). *Special Educational Needs: Support in England*. [online] Parliament.uk. Available at:
<https://researchbriefings.files.parliament.uk/documents/SN07020/SN07020.pdf>

Malkani, R. (2021). Investigating the opportunities provided for young people with special education needs and disabilities (SEND) to prepare for adulthood in a city in England. *Support for Learning*, 36(2), 238-257.

McCrary, Justin. (2008) Manipulation of the running variable in the regression discontinuity design: A density test, *Journal of Econometrics*. 142(2): 698-714.
<http://dx.doi.org/10.1016/j.jeconom.2007.05.005>

Migliore, A., Timmons, J., Butterworth, J., & Lugas, J. (2012). Predictors of employment and postsecondary education of youth with autism. *Rehabilitation Counseling Bulletin*, 55(3), 176-184.

Misik, L., & Barnhardt, E. (2024). Socioeconomic status remains largest predictor of special education provision in the UK. *Evidence-Based Nursing*.

OECD (2022), *Disability, Work and Inclusion: Mainstreaming in All Policies and Practices*, OECD Publishing, Paris. <https://doi.org/10.1787/1eaa5e9c-en>.

Abreu, L. (2024) Autism: overview of policy and services. House of Commons Library Briefing Paper, (7172).

Peckham, T., Fujishiro, K., Hajat, A., Flaherty, B. P., & Seixas, N. (2019). Evaluating employment quality as a determinant of health in a changing labor market. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 5(4), 258-281.

Sibieta, L. and Snape, D. (2024). *Spending on special educational needs in England: something has to change IFS Report R341*. [online] Available at:

<https://ifs.org.uk/sites/default/files/2024-12/Spending-on-special-educational-needs-in-England.pdf>

Stacey, E., & Dewey, J. (2024). Perspectives of young people with SEND on a supported internship programme in the UK. *British Journal of Special Education*

Tyllianakis, E., Rigas, N., Kounetas, K. (2025). The impact of announcing a payment assistance program on UK household water consumption: A regression discontinuity analysis. *Utilities Policy*, 96, 101981.

UK Parliament Report. (2021) Disabled people in employment. London: House of Commons Library.

UK Parliament Report (2024). *Solving the SEND Crisis*. [online] Available at: <https://publications.parliament.uk/pa/cm5901/cmselect/cmeduc/492/report.html>.

Wehman, P., Schall, C., McDonough, J., Graham, C., Brooke, V., Riehle, E., Brooke, A., Ham, W., Lau, S., Allen, J., & Avellone, L. (2016). Effects of an employer-based intervention on employment outcomes for youth with significant support needs due to autism. *Autism*, 21 (3), 276–290. <https://doi.org/10.1177/>

Wehman, P., et al., (2020). *Competitive employment for transition-aged youth with significant impact from autism: A multi-site randomized clinical trial*. Available [here](#).

Winefield, A. H., & Tiggemann, M. (1990). Employment status and psychological well-being: A longitudinal study. *Journal of Applied Psychology*, 75(4), 455–459. <https://doi.org/10.1037/0021-9010.75.4.455>

Wilson, S. (2021, September 6). The MICE Algorithm. Retrieved from <https://cran.r-project.org/web/packages/miceRanger/vignettes/miceAlgorithm.html>

Youth Futures Foundation (2026). What is 'good work'? Retrieved from <https://youthfuturesfoundation.org/about-us/what-is-good-work/>

7. About the authors

The Policy Institute at King's College London works to solve society's challenges with evidence and expertise. Part of the Faculty of Social Science and Public Policy, we combine the rigour of academia with the agility of a consultancy and the connectedness of a think tank. The research was conducted by researchers in the Policy Institute's Evidence Development and Incubation Team (EDIT) who provides evaluation services for a diverse range of clients across the public and third sector. EDIT was founded to address the growing need for high-quality causal evidence that shows "what works" in public services, education, youth employment, homelessness, and other policy area.

Irene Soriano Redondo is a Research Associate at the Policy Institute. She is an experienced qualitative researcher and evaluator with experience in conducting research with children and young people, as well as social inequalities.

Domenica Avila-Luna was a Research Associate at the Policy Institute. She is an experienced quantitative researcher, with experience conducting experimental and quasi-experimental evaluations.

Stephen Hunsaker was a Research Associate at the Policy Institute. He specialises in quantitative analysis, supporting the design and delivery of feasibility studies and evaluations.

Dr Paolo Morini is a Research Fellow at the Policy Institute. He is an experienced quantitative researcher, with experience in public opinion and public engagement.

Beti Baraki is a Research Fellow at the Policy Institute. She specialises in policy evaluation and analysis, qualitative and participatory research across different areas, including inequality and wellbeing.

Johnny Runge is a Senior Research Fellow at the Policy Institute. He is an experienced qualitative researcher, including on employment, education, and inequalities.

Dr Susannah Hume is Director of Evaluation at the Policy Institute. She served as the establishing Director for the Centre for Transforming Access and Student Outcomes (TASO) and is on the Cabinet Office's Evaluation and Trials Advisory Panel and the Youth Futures Evaluation Advisory Group. Dr Hume was the Principal Investigator on this feasibility study.

Appendix A: Rubric Sheet



Overall Combined Score

Student name _____ Date Completed _____ Completed By _____

TOTAL SCORE: (out of a possible 100): _____

Criteria	1	2	3	4	5	Source
Age and School Status	Under 18 or over 24 years of age	Two or more years of school left, lacks maturity and working towards employability	Needs one or two qualifications for graduation and can be fulfilled through Project SEARCH	Has two years of school eligibility/ agreed will be the last year/ has the goal of employment	In last year of school, has all the qualifications necessary for graduation	Application Form/school knowledge
Commitment to Community Employment	Is unsure of interest in employment	Is unsure of interest in employment, but parent is supportive and encouraging	Commitment to work shown but has restrictions: location inappropriate work goal, type of work,	One member of the team, student, or family non-committal to the goal of employment	Student and family are committed, supportive and appropriate; they will support the work goal and assist with transportation	Interview

Attendance	10+ unexcused absences	10+ excused and unexcused absences/ unpunctual in past school year	5-10 excused and unexcused absences/ unpunctual in past school year	1-5 excused absences or unpunctual within the past school year	No absences or punctuality issues within the past school year	Application Form/school knowledge
Independent Daily Living and Self Care Skills	Has very poor or no independent daily living and self-care skills. Reliance on parents and staff for some basic needs to be met	Has not been exposed to any daily living skills training but displays some skills in these areas and in self-care.	Has participated in limited or informal training for daily living and self-care skills. Can demonstrate minimal skills. Accommodations could support.	Basic proficiency in daily living skills and self-care skills. Self-care skills include toileting, feeding, taking medication, bathing, etc.	Student demonstrates daily living and self-care skills such as cooking, sleeping, budgeting, handling money, and is also able to take care of self-care needs independently	School knowledge
Appearance and Professional Presentation	Does not possess any personal hygiene skills and clothes are not neat and clean	Needs assistance in making sure clean clothes are worn daily; appearance may vary each day	Wears neat and clean clothing and has appropriate grooming on most days	Neat, clean, well-groomed but makes inappropriate clothing choices for work place, dress code and weather.	Possesses good personal hygiene skills. Always arrives to school neat/clean, according to dress code and weather	School knowledge and interview
Transportation	Not eligible and/or has not been trained in any independent travel. Family cannot provide transportation	Family is willing to provide on-going transportation to the Project SEARCH and/or work site	Eligible for transportation from outside resources and/or has access to public transportation	Can utilize public transportation including a door-to-door system with basic skills	Can utilize public transportation and/or a door-to-door system as well as other transportation options. A Project SEARCH applicant rarely drives.	Interview
Appropriate Social and Behavior Skills	Frequently displays inappropriate	Periodically displays inappropriate	Appropriate in the presence of adult	Displays appropriate social and behavior	Displays appropriate social and behavior	Assessment Activity

	social and behavior skills	social and behavior skills	supervision but is not independent	skills in most situations	skills in all (or nearly all) situations	
Interpersonal Communication	No grasp of interpersonal relationships	Uses appropriate body language (smiles, waves, etc.) but doesn't engage in appropriate conversation	Engages in conversation when prompted	Engages in conversation independently, but the topic is inappropriate	Uses appropriate tone of voice, body language and conversation topics	Assessment Activity and Interview
Verbal Communication	Has no way of clearly communicating with others	Not fluent or easily understood even with assistive technology or support	Can be understood with one to two repetitions or when asked to speak more clearly	Uses a voice box or assistive technology to communicate and is understood using these tools	Has the ability to communicate with all others and be understood easily	Assessment Activity and Interview
Problem Solving and Conflict resolution	Has no independent problem solving and conflict resolution skills	Has difficulty in problem solving and conflict resolution skills	Has demonstrated capacity to expand problem solving and conflict resolution skills	Possesses good problem-solving skills	Possesses good problem-solving skills and initiates problem solving independently	Assessment Activity
Physical Ability ■ Mobility ■ Stamina	Has limited physical abilities as well as limited capacity to take care of own personal needs	Has the mobility and stamina to perform some of the tasks including personal needs with limitations	Has low but improving mobility and stamina and ability to take care of own personal self-care needs	Has mobility and stamina to perform tasks with accommodations for mobility. Has strategies to take care of personal needs with limited assistance	Has the ability, mobility and stamina to perform all tasks including self-care independently and successfully	School knowledge and interview
	Seldom gets work finished in	Is methodical which affects pace,	Can achieve appropriate work	Pace and quality of work are mildly	Can achieve both quality and quantity of	Assessment activity

Pace and Work Quality	allotted time or by deadline	productivity, and output	pace but quality suffers, or work quality is ok, but quantity is affected	deficient but improving and work is neat	work, maintains neatness/organization and completes work according to deadlines	
Employability Skills	Has not been exposed to any employability training	Has participated in limited or informal employability training	Has had one year or a consistent class devoted to employability skills	Has had more than one year of employability skills training	Demonstrates independent living skills. Has 2 or more years of employability training	Application form and interview
Prior Work Experience	Has no prior work experience	Has had one or more in-school work experience	Has volunteer or paid enclave work experience	Has had one or less years of competitive work experience	Has had more than one year or multiple work experiences	Interview
Academic Skills	Cannot read nor do any simple math computations	Has some academic skills such as rote counting. Can file using two to three digits with numbers or letters	Can read simple functional information and can perform simple math computations (with or without a calculator)	Can read /understand simple material, tell time on clock or analog watch to five minutes/count money/make change	All academic skills are at or above GCSE level	School knowledge and assessment
Computer Skills	Has no computer skills	Has basic knowledge of keyboard and keyboard functions	Can access internet, utilize search engines, and uses for entertainment	Can utilize some Microsoft products at a beginner level and emails appropriately	Can utilize Microsoft, save, edit and retrieve documents with basic proficiency and emails appropriately	Assessment activity
Values	Lacks pride in work, refuses to participate, disrespectful	Will work with many prompts and corrections	Compliant with requests and can complete task list with prompts	Follows requests willingly, wants to do work well, polite with co-workers/managers	Honest, takes pride in work, excited to participate, exhibit core values of organization	Assessment activity

Self-Advocacy & Initiative	Sits down or wanders after given tasks	Asks for more work with prompting; rarely initiates tasks independently	Asks for new task independently	Asks for additional work after tasks are completed; utilizes accommodations to improve work	Understands own strengths and challenges, requests accommodations; problems solves on own	Assessment activity
Focus & Following Directions	Constantly distracted by movement, visuals and noise	Can stay on task for short periods of time	Can focus on work for 30 minutes or longer	Can focus on task for at least an hour with no breaks	Can focus on and finish task with multiple distractions	Assessment Activity
Safety	Can't recognize unsafe situations, runs away, and/or has some physical aggression	Needs close supervision in many situations to keep himself and others safe	Is safe at least half of the time with prompts	Can recognize most safe situations with occasional prompts	Is safe in all situations and respectful to all	Interview

Overall Assessment Rubric Score:
Skills Assessments and Interview Scores:
TOTAL SCORE:

Appendix B: Survey Script

Survey Script

The survey will be developed in Qualtrics. All associates and DFNPS personnel supporting the data collection should register their answers in a Qualtrics form that will be provided by the research team. Below, we provide a survey script:

Note for site assessors: Ahead of the first survey, please make sure you have reviewed the briefing document provided by King's. The document outlines key instructions that should be followed when conducting the surveys with young people. As indicated in the briefing document, please use the printed cards to support participants that might need visual support to answer the survey questions. All notes to for you are in grey (do not read these to participants).

Key steps before starting the survey:

1. Make sure you have a laptop to input the answers in the Qualtrics online form, and printed cards with emojis/pictures to help participants answer the questions you will ask them
2. Explain to the participant what the research is about (please use the text provided in the briefing document). Please emphasise that the survey is confidential and voluntary, and that there are no right or wrong answers.
3. Obtain written consent from participant (please follow instructions provided in the briefing document). Please ensure participants provide explicit consent to all statements in the consent form.

Site assessor to start the survey and read out loud:

To start, I would like to collect some information about you.

[General information – ASK ALL]

1. What is your full name? [Open box]
2. Which college do you go to? [Drop down list]
3. What is your date of birth? [Format dd/mm/yyyy]
4. What is your postcode? [Open box]
5. What is your sex? A question about gender identity will follow later on in the questionnaire [Single choice]

- male
 - female
6. Is the gender you identify with the same as your sex registered at birth? [Single choice]
- Yes
 - No

[ONLY IF THEY SELECTED NO]

6.1 What is your gender identity? [Open box]

7. Have you had paid work before? [Single choice]
- No
 - Yes

[ASK IF YES TO Q6]

- What job was it?

[Confidence – ASK ALL]

Now, I would like to know how much confident you feel doing certain things.

[Please show cards with emojis to participants]

1. Do you feel confident in expressing your opinion? [Single choice, use scale 1 to 3, or 99]
- a. No [=1]
 - b. Sometimes [=2]
 - c. Yes [=3]
 - d. Don't know [=99]

2. Do you feel confident in working or studying with peers? *[Single choice, use scale 1 to 3, or 99]*
3. Do you feel confident helping a new person feel included in your group? *[Single choice, use scale 1 to 3, or 99]*
4. Do you feel confident sharing a story about something you enjoyed with new people? *[Single choice, use scale 1 to 3, or 99]*

[Social and communication skills – ASK ALL]

Great!

Now, I will read you a few questions about communication, please let us know how you feel most of the time. *[Please show cards with emojis to participants]*

[Communication]

1. Do you feel interested in what peers are saying *[Single choice, use scale 1 to 3, or 99]*

[Cooperation]

2. Do you follow directions? *[Single choice, use scale 1 to 3, or 99]*

[Assertion] or self-advocacy or self-expression

3. Do you stand up for yourself when treated unfairly? *[Single choice, use scale 1 to 3, or 99]*

[Responsibility]

4. Do you take responsibility when you have done something wrong? *[Single choice, use scale 1 to 3, or 99]*

[Empathy]

5. Do you try to understand how your friends feel when they are angry, upset or sad. *[Single choice, use scale 1 to 3, or 99]*

[Engagement] relationships and engagement skills

6. Do you enjoy spending time with peers *[Single choice, use scale 1 to 3, or 99]*

[Health and wellbeing – ASK ALL]

Perfect, now I would like to know about your health.

2. We would like to know how good or bad your health is TODAY.

The scale is numbered from 0 to 100.

100 means the best health you can imagine.

0 means the worst health you can imagine.

1. Can you tell when you are feeling unwell or in pain? *[Single choice, use scale 1 to 3, or 99]*
2. Do you know what helps when you feel worried or upset? *[Single choice, use scale 1 to 3, or 99]*
1. Do you feel excited about your future? *[Single choice, use scale 1 to 3, or 99]*
2. Are you feeling calm today? *[Single choice, use scale 1 to 3, or 99]*
3. Can you think clearly when you have a problem? *[Single choice, use scale 1 to 3, or 99]*
4. Did you make decisions for yourself today? *[Single choice, use scale 1 to 3, or 99]*

[Financial independency – ASK ALL]

We are about to finish! The last questions are about your money and how you manage it.

I will read you three questions, and I'd like you to tell me if they are true for you or not. *[Please show cards with emojis to participants]*

1. Do you know how much money you have to spend? *[Single choice, use scale 1 to 3, or 99]*
2. Do you end up spending more than you planned at the store? *[Single choice, use scale 1 to 3, or 99]*

3. Do you get to choose how to spend your money? *[Single choice, use scale 1 to 3, or 99]*

That's all! Thank you for taking part in this survey!

Appendix C: Causal Mechanisms

DFNPS' Theory of Change presents the anticipated outcomes of the programme to young people that have participated in supported internships, their families or carers, employers and host businesses, educators and colleges, LAs, the state and wider society.

Young people in supported internships

As reflected in the diagram, in the short-term, young people taking part in a supported internship are expected to experience an increase in their confidence, self-esteem and aspirations as they take their first steps into a formal, supported work environment. Research shows that supported internships that provide autonomy in task development was associated with an increase in participants' self-esteem (Folguera, 2025; Romualdez et al., 2020). Their work with the employability tutor and job coach is also expected to support the achievement of these outcomes while they are in the supported internship and beyond, as they learn new skills that can be applied after the supported internship. While the employability tutor and job coach play a key role in supporting interns during their internships, the sustainability of these outcomes beyond the internship is reinforced through ongoing interactions with other outcomes – such as improved health and wellbeing – as illustrated in the ToC.

Their participation is also expected to lead to improved practical and transferrable work skills, as they gain exposure to new tasks and responsibilities in a work environment through their rotations. This supported experience and the accessibility of employability tutors and job coaches is expected to also increase their emotional independence, and support the development of social, emotional, communication and self-advocacy skills through their interactions with work colleagues, coaches and tutors. There is evidence that shows that supported internships support individuals with SEND to develop social skills and increase social inclusion (Hanson et al., 2021).

Alongside these outcomes, the programme is also expected to improve participants' health and wellbeing while they are part of the supported internship, but also beyond the duration of the programme. The improved wellbeing is an anticipated result from participants improved mental and physical health, as interns positively interact with new colleagues, perform physical tasks, commute to work, and are overall physically active during their job (Ashworth et al., 2025; Hanson et al., 2021; Viner et al., 2018; Ellenkamp et al., 2016).

In the medium term (after the internship is completed), participants are expected to transition into paid employment or apprenticeships, which would improve their earnings and financial independence (Lindstrom et al., 2013). The programme aims to support this by facilitating the development of transferable skills that can be then applied to job search and to retain current roles. The work interns do with employability tutors and job coaches is also expected to lead to improved interview skills, which are key to succeed in recruitment processes.

Research focusing on interns of supported internships transition into employment highlight the following factors as the most significant contributors to a successful transition: inclusive work environments (Beyer & Kaehne, 2008), on the job training (Beyer & Kaehne, 2008) and sufficient experiences to allow the interns to develop and identify their preferences to help them make informed decisions (Kaehne, 2016).

In the long-term, all these outcomes are expected to allow participants to transition into sustainable, paid employment.

Families and carers

Families and carers of interns who are participating in supported internships are expected to experience a range of benefits. The increase in interns' emotional independence, communication and self-advocacy skills, is expected to reduce caring responsibilities of families and carers in the short-term. These changes, together with the improved financial independence of interns in the medium-term, can also contribute to improved financial situation of families, as some parents and carers might (re-)enter the workforce in the long run.

Research suggests that individuals with SEND – including those with LD and/or autism - typically have higher expectations of themselves and their capabilities compared to their families and employers (Holwerda et al., 2015; Ebuonyi et al., 2020). Demonstrating their potential through their supported internship and developing social and self-advocacy skills is expected to increase family pride as well as reduce their worries.

Overall, as interns develop their employability skills and secure new, paid, sustainable job opportunities, families are anticipated to improve their optimism about the future. The increase in young people' financial independence – which can contribute to cover personal expenses or bills – is anticipated to reduce care giver fatigue. In turn, this reduction in fatigue can support family members to return to employment.

Employers and host businesses

Bringing on board DFNPS interns is expected to bring a range of benefits for host employers and businesses more widely. In the short term, hosting interns is anticipated to allow employers to upskill their existing teams, as employees learn how to support interns and take on roles such as mentoring that contribute to team development. At the same time, interns gaining hands-on experience and developing key skills to perform their tasks well is expected to reduce skills gaps, improving the overall performance of the host business.

Changes in perceptions and reduced stigma of what people with LD and/or autism can do and achieve is also an anticipated short-term outcome. By hosting interns with LD and/or autism that fulfil different tasks and take on different responsibilities, staff members increase their exposure with young people with LD and/or autism and recognise existing biases that tend to exclude talented individuals with LD and/or autism.

In the medium term, retaining interns with LD and/or autism after their internships increases workforce diversity. In turn, this is expected to lead to long-term outcomes such as increased inclusivity of recruitment practices that support individuals with SEND more broadly (e.g., such as creating adjustments to interview processes), improved general retention rates and a more inclusive team culture, as inclusion of employees with LD and/or autism and facing other additional barriers is normalised within organisations.

Educators and colleges

DFNPS aims to achieve positive outcomes for educators and colleges that work with young people with LD and/or autism. By signposting young people with LD and/or autism to supported internships, colleges expand their range of post-mainstream education opportunities enhancing their ability to offer meaningful pathways that support higher success rates and positive long-term outcomes for young people. In the long run, this can lead to improved Ofsted rankings, as the colleges would be able to show strong post-16 progression routes and employability outcomes for young people with LD and/or autism. Additionally, this can enhance the college's reputation and attract more young people. This improvement comes from better advertising of successful internship placements, a strong commitment to inclusive learning environments, and potentially stronger partnerships with employers. These collaborations create more work experience opportunities, benefiting both young people and the colleges.

Local authorities, the State and wider society

In the medium and long term, the programme is expected to lead to wider benefits to LAs and the State. The employment of young people with LD and/or autism following a supported internship is expected to lead to social and economic value, which in turn can lead to reduced public expenditure. This would result from a reduction in the number of people receiving benefits/amount of benefit received (and an increase in taxpayers/amount of tax paid) and reduced use of healthcare services due to individuals' improved health (Lindsay et al., 2018; Thomas & Ellis, 2013).

The employment of people with LD and/or autism in different sectors is also expected to contribute to reducing local skills shortage within specific sectors where DFNPS operates. Additionally, interns gain transferrable skills that can support broader employability. Alongside this, improved outcomes for young people with LD and/or autism can enhance LA's reputation by demonstrating commitment to inclusive employment.

More widely, in the long term the programme is expected to contribute to a change in the way society perceives what people with SEND, including those with LD and/or autism can do and achieve, as they are more visible in workplaces, and take up new roles and responsibilities within organisations.

Key assumptions and risks

The key assumptions identified during the workshop that underpin this ToC are:

- The programme recruits the right young people. That is, those that are ready, willing and would benefit the most from taking part in a supported internship. This is important because if the recruited participants are not suitable, they may drop out and fail to complete the programme.
- There are suitable vacancies in participants' local area and there is a suitable pool of young people to fill the host businesses' programmes
- Host businesses provide real opportunities to interns and ensure they rotate within the business, allowing interns to perform different tasks and learn different skills. This directly affects skills development for young people and work readiness.
- Host businesses provide supportive environments where interns can learn new skills, and positive relationships between tutors, job coaches and interns are established to ensure positive outcomes are achieved. Positive relationships is also central to fostering sense of belonging leading to better wellbeing and in turn retention.
- Post engaging with young people with LD and/or autism, host businesses are more willing to create change within their organisation (e.g., changes to their recruitment practices, if needed) to improve workplace diversity.

Tutors and job coaches are skilled in delivering the programme, and the programme model is followed by participating stakeholders, ensuring that young people with LD are receiving consistent support across the sites.

- An increased visibility of young people with LD in the workplace reduces stigma around these population and allows people to recognise existing biases that tend to exclude talented individuals with SEND.
- Labour market are stable enough to provide placements for young people, without which, host businesses may not have the capacity to provide internships.
- Follow-on support (post-internship) is available across sites to ensure young people are supported to transition from internships into jobs and outcomes are sustained.

The key risks that could interfere with the programme's performance are:

- There are no or limited jobs available for participants
- Participants don't complete the programme
- Host businesses face financial difficulties and instability
- High staff turnover leads to challenges replacing educators and tutors that leave their job affecting programme delivery.
- Follow-on support not available for those who have or have not secured employment in other organisations outside their host business

References

Ashworth, M. C., Heasman, B., Crane, L., & Remington, A. (2025). Tracking the long-term outcomes of a supported employment internship programme for autistic adults without a learning disability. *Neurodiversity*, 3, 27546330241311472.

Beyer, S., Kaehne, A., Grey, J., Sheppard, K., & Meek, A. (2008). What Works? Transition to employment for young people with learning disabilities. *Final Report*.

Ebuenyi, I. D., van der Ham, A. J., Bunders-Aelen, J. F., & Regeer, B. J. (2020). Expectations management; employer perspectives on opportunities for improved employment of persons with mental disabilities in Kenya. *Disability and rehabilitation*, 42(12), 1687-1696

Ellenkamp, J. J., Brouwers, E. P., Embregts, P. J., Joosen, M. C., & van Weeghel, J. (2016). Work environment-related factors in obtaining and maintaining work in a competitive employment setting for employees with intellectual disabilities: A systematic review. *Journal of occupational rehabilitation*, 26, 56-69.

Folguera Bellmunt, C. (2025). Supported internships: vocational training and labor inclusion for students with intellectual and developmental disabilities. *Equality, Diversity and Inclusion: An International Journal*.

Hanson, J., Robinson, D., & Codina, G. (2021). Supported internships as a vehicle for social inclusion. *British Journal of Learning Disabilities*, 49(4), 433-444.

Holwerda, A., Brouwer, S., de Boer, M. R., Groothoff, J. W., & van der Klink, J. J. (2015). Expectations from different perspectives on future work outcome of young people with intellectual and developmental disabilities. *Journal of occupational rehabilitation*, 25, 96-104.

Lindsay, S., Cagliostro, E., Albarico, M., Mortaji, N., & Karon, L. (2018). A systematic review of the benefits of hiring people with disabilities. *Journal of occupational rehabilitation*, 28, 634-655.

Lindstrom, L., Kahn, L. G., & Lindsey, H. (2013). Navigating the early career years: Barriers and strategies for young people with disabilities. *Journal of Vocational Rehabilitation*, 39(1), 1-12.

Romualdez, A. M., Yirrell, K., & Remington, A. (2020). Exploring participants' views on a supported work internship program for autistic and learning disabled young people. *International Journal of Disability Management*, 15, e3.

Thomas, K. C., & Ellis, A. R. (2013). Patterns of healthcare use and employment among people with disabilities. *Disability and health journal*, 6(2), 133-140.

Viner, R., Hargreaves, D., Varnes, L., Heys, M., & Viner, R. (2018). The social determinants of young people's health. *Health*.

Appendix D: Observation guide

This guide is to be used for the observation of activities that form the DFN Project SEARCH (DFNPS) programme. As a variety of activities will be observed, the prompts included below are intentionally broad and should not be viewed as an exhaustive list.

All the activities will involve interactions between DFNPS staff, a team of assessors, and programme participants. The working relationship between participants and the staff is central to the intervention and should be a key focus of the observations.

Observations must not interfere with the assessments or verbal interviews that take place in sites, to avoid disrupting participants' performance. For this, ahead of the observation (or on the day of the observation) please check with assessors or staff which are the specific areas or activities where your presence is most adequate, and adjust accordingly.

Please minimise your visibility (e.g., by sitting or standing in an area that is out of the direct line of sight of participants) and defer questions to avoid interrupting. To ensure observations are as smooth as possible, periodically check with assessors to ensure our presence is not causing any unintended disruption.

Prompts:

The below prompts should be considered in relation to both the participant(s) and the member(s) of staff.

- **Environment** (including the building, set up of the room, equipment in the room, furniture, accessibility etc.)
- **Who's in the room** (descriptions of role, relationship, what they're wearing etc)
- Topic of conversation (and who initiates it)
- **Body language** (level of comfort shown, familiarity etc.)
- **Tone** (happy, relaxed, guarded etc.)
- **Type of language** (formal or casual, choice of words etc.)
- **Flow of conversation** (who is leading, stilted or flowing etc.)
- **Materials used** (computer, visual aids, paperwork etc.)
- **Attitude** (positivity, apathy etc.)
- **Agency** (roles in determining activities)

Site Name: Day:	Notes
Brief description of the environment (e.g., venue, size of the room, noise levels, accessibility, number of people in the venue)	
Participants' arrival at the venue (e.g., interaction with programme staff, body language, gestures or expressions of young people at arrival)	
Pre-assessment (e.g., introduction to the programme and to the assessment)	

activities, engagement, gestures or expressions, materials used)	
Assessment interviews (e.g., set up, who is involved, body language, attitude)	
Assessment tasks (e.g., type of task, set up, materials used, participants' attitude, agency and engagement)	
End of the assessment (e.g., what happens at the end of the assessment, participants' leaving the venue)	

Staff's assessment of candidates (e.g., use of rubric scores, materials used, type of discussion with other staff to assess candidates)	
Other reflections of observer (e.g. any observed differences between candidates, ethnic diversity, disability type, any strategies used by assessor to make assessment accessible)	

Appendix E: Feasibility Assessment

To assess the feasibility of the different impact designs, we used Progression Criteria to facilitate Youth Futures decision about whether it is possible to progress to a full evaluation, and which is the most suitable design to do so. The criteria, initially agreed as part of the study's feasibility plan, was further refined to accurately capture different elements that determine the overall feasibility of each impact evaluation method.

1. Robustness
2. Practical feasibility
3. Acceptability of the design
4. Working relationship with the delivery organisation and other relevant stakeholders
5. Ethical considerations

These Progression Criteria used a traffic light system representing a 'Readiness Indicator', consistent with the following:

Green	The criteria analysed is good to go to a full-scale evaluation
Yellow	There are reasonable improvements to be made to the criteria to conduct a full-scale evaluation
Red	The criteria analysed is not ready to undertake a full-scale evaluation, and the changes required are considerable or not doable in an appropriate timeline

The criteria are explained below in more detail.

Robustness of the proposed design

This criterion explores the extent to which the impact methodology produces robust results, taking into account key enablers and barriers. This criterion mainly considers if the currently available sample of young people in DFNP is sufficient for conducting an RCT or QED. This also involves computing the size of a potential control group when exploring an RCT, or the size of the comparator pool if exploring a QED.

As part of this, we also conducted initial sample size calculations to set high and low scenarios of sample sizes for a range of minimum detectable effect sizes. This information provides valuable insights into whether it is necessary to extend the recruitment to reach better sample sizes and to discuss what would be a desired/expected effect size in the context of this evaluation, and the sector more broadly.

This progression criterion will be assessed for each design under consideration as follows:

The available sample size is large enough that allows to capture a small Minimum Detectable Effect Size (MDES) of 1.43 *odd ratios* and allows for generalisability of the results. This means that the design can capture an effect of the treatment group being at least 1.43 times more likely to access a sustainable job, than the control/comparator group.

The sample size allows to detect a moderate MDES of between 1.43 and 2.47 *odd ratios*, which could be sufficient to provide meaningful insights, but will require potentially increase the sample size and enhance statistical power further.

The sample size allows to detect a MDES of more than 2.47 *odd ratios* and increasing the sample size will not be possible as part of the full-impact evaluation.

Practical feasibility

This aspect refers to whether the current delivery of the programme shows promise for the implementation of an impact evaluation. There are several factors to be considered including:

- Referral pathway: How are young people referred into DFNPS? Are there any filtering happening along the referral journey that could be exploited to construct a comparator group?
- Eligibility criteria: What are the characteristics considered when placing a young person in a DFNPS supported internship?
- Consent and data sharing requirements
- Consistency of programme delivery and dosage
- Randomisation feasibility and risk of spillover effects/contamination
- Exogeneity conditions to exploit to facilitate a QED
- Identification of a credible comparator group (QED)
- Outcomes measurability

- Availability of required data, both through administrative records and surveys.

Based on these factors, the progression criteria was assessed as follows, separately for each design under consideration:

The design is practically possible, as each of the factors above are either met, or there is a clear route to meet them, and no major challenges identified. This indicates a high level of feasibility and minimal risks in conducting the study. Special attention will be given to randomisation feasibility, and identification of a credible comparator group, and data availability.	The design may be practically possible, but there will be challenges to overcome and some significant risks. Either the randomisation feasibility factor or the identification of a credible comparator group is met, but other factors require additional effort and some resources to address the challenges and ensure a successful study.	The design is not practically possible, or there would be significant risk the approach would fail. There is no plausible randomisation option and there is no credible comparator group available. Other factors also indicate substantial changes to the study design or important resources would be needed to address the challenges and risks.
---	---	--

Acceptability of the design

This section refers to whether the stakeholders involved (DFNPS staff and colleges) are supportive of an impact evaluation generally, and of the specific design under consideration. To evaluate this, we conducted a workshop with key personnel from DFNPS where to explore the possibilities of conducting an RCT or a QED. In this workshop, we assessed whether different stakeholders felt comfortable with the approach, whether it aligns with the organisation core values, and to what extent the design represents a burden when implemented.

This progression criterion was assessed for each design under consideration as follows:

All stakeholders involved including DFNPS, site assessors and colleges are	DFNPS is supportive of exploring the design(s) while some other stakeholders hold less	DFNPS is not supportive of exploring the design(s) being conducted, and
---	---	--

open to exploring the design(s) being conducted

supportive views on the evaluation; or DFNPS is not very supportive of the design but there is a clear path with activities to gain their buy-in for the research design.

there are paths to gain that support.

Working relationship with the delivery organisation

Throughout the evaluation, we assessed whether there is a constructive and open working relationship between the evaluator and DFNPS. We considered how well the evaluators and DFNPS can work together and have mutual trust in each other throughout the process. We also considered aspects such as open communication, joint buy-in to conducting an impact evaluation, understanding of the importance of the evaluation and willingness to accept the evaluation outcome, if it doesn't show promise. We also considered the capacity of the organisation to provide the necessary support (time and resources) for a full impact evaluation going forward, or what could be done to guarantee that capacity moving forward. We also assessed the DFNPS willingness to adapt/change their process in a timely manner to support the evaluation. Similarly, we explored how DFNPS manages their relationship with colleges and sites.

This progression criterion was assessed as follows:

The evaluator (KCL) and DFNPS have a strong working relationship and are committed to the research, and there is confirmation of availability of sufficient resources to support a full experimental or QED design. Additionally, DFNPS and the evaluator (when relevant) have a strong working relationship with a number of sites and colleges that is big enough to deliver sufficient

Most stakeholders, but not all express commitment to the research and have a positive working relationship, and there is availability of some resources to support full experimental or QED design, but additional resources may be needed.

The relationship of the evaluator and DFNPS is not strong enough to conduct the design, including low capacity/resources. Additionally, the relationships between the delivery partner and colleges/sites are weak and there is not enough commitment to guarantee their contribution to the research.

numbers for a powered full-impact design.

Ethical considerations

Throughout the feasibility phase, the research team assessed any risks that participants may face when engaging with a full-scale evaluation. The team also assessed which mitigation strategies and safeguarding procedures are needed according to the level of vulnerability of the participants to ensure they can safely take part on the project. This criterion therefore assesses whether in the context of DFNPS it is possible to propose an evaluation design that ensures adequate ethical standards. Possible sensitive points from an ethics perspective are:

- **RCT design.** If this is chosen to be the most appropriate approach for the evaluation, we will provide evidence to the ethics committee on how the delivery organisations are comfortable with the approach, and how the design does not pose a threat to the participants or the control group. We will explore how feasible is to conduct either an RCT with oversubscription or a waitlist design, or whether it is possible to exploit any other source of exogeneity in intern selection. We will also provide evidence on how risks can be mitigated throughout the course of the research.
- **Consent.** We will assess the best way to involve participants in the research and ensuring that they are fully aware of what their participation entails. This also involves assessing how easy-read consent forms and information sheets have worked during the feasibility phase.
- **Instruments.** We will consider whether any survey materials are potentially distressing for the participants, and how to mitigate/avoid this.

This progression criterion will be assessed as follows:

There is a clear path to obtain ethical clearance, along with the necessary information to obtain approval while mitigating risks for young people

More information is required in order to submit an ethics application for clearance, and that information can be easily obtained

There is no clear path to obtain ethical clearance as it is not possible to mitigate the risks derived from the research for young people

Appendix F: Other designs considered

As part of our previous Recommendations Paper (not published), we explored three additional designs and identified them as unsuitable for a full impact evaluation of DFNPS. This section briefly outlines the reasons why we do not recommend these methodologies for a future evaluation of the programme.

Randomised Controlled Trial at college level

A college-level RCT was explored but discarded due to several critical issues. In this design, randomisation to DFNPS would take place at college level, with colleges in the treatment group able to offer DFNPS supported internships, and those randomised to the control group not able to participate in DFNPS. The impact of DFNPS would be estimated by comparing the outcomes of students enrolled in treatment or control colleges.

We advised DFNPS to not proceed with this design on ethical, methodological, acceptability, and working relationship grounds. Ethically, withholding provision of DFNPS to students in control group colleges would not be acceptable as we identified alternatives that avoid this issue. Additionally, withholding provision would also affect DFNPS's relationships with colleges who are randomised to the control group, which makes the design unacceptable for the DFNPS leadership. There are also practical concerns around the need for incentives to engage control college which can increase costs to run the RCT considerably.

Therefore, while the college-level RCT would offer robust results, the practical, ethical, and relational challenges make it an inadequate approach for this context, and other designs are more suitable for the evaluation of DFNPS.

Matched comparator analysis

In a matched comparator design, the impact of DFNPS would be estimated by comparing participants who took part in DFNPS with comparable participants who didn't. The comparison group would not be identified through randomisation, but by using matching techniques which specifically select observations based on observable characteristics such as ethnicity, gender, age, and past experiences, so that they match the characteristics of the treated sample. The impact of DFNPS would be estimated by

comparing the outcomes of treated students with those from the matched comparator group.

While we identified that, in theory, this would be a feasible evaluation approach, we have crucial concerns about several limitations that are unlikely to be overcome in an evaluation of DFNPS. Firstly, the evaluation would face key barriers when seeking to access data on employment outcomes for comparator colleges. Specifically, as of 2024/25, the DfE has stopped requiring colleges to report employment outcomes post-course into the ILR. While other data sources could be considered instead, these are unlikely to work. And if data is instead accessed through comparator colleges, they would need to be recruited and engaged in the study, adding additional complexity.

Secondly, questions would arise on what programmes are offered by colleges of matched comparator individuals, as even if the comparator group does not participate in DFNPS, they may still receive alternative forms of support (e.g., local supported internships within their colleges, employability training). This would result in comparisons between DFNPS and multiple forms of control points, some of which might be very similar to DFNPS, reducing the possibility of identifying a significant effect.

The DFNPS leaderships also expressed low support for this design, especially as it queried whether this design could capture the programme's true effect given the existence of comparable programmes in the comparator group. Overall, a matched comparator approach would be logistically too complex and challenging, especially compared to other more robust designs we presented above.

Individual-level Difference-in-Differences

In an individual-level DiD approach, the impact of DFNPS would be estimated by comparing the change in employment status in time for people who took part in DFNPS and those that didn't. Identifying a comparator, much like we discussed for the matched comparator approach above, would prove complex in the current data environment, with no information available on young people's employment outcomes, and with no information about business provisions for comparator observations.

For the former, the data on employment and education outcomes are crucial to test for the existence of parallel trends, an assumption for the robust identification of DiD effects, but given the currently available information, we would at best have suggestive evidence of the design meeting the assumption. For the latter, as with the matched comparator case, the DiD would not capture the effect of DFNPS provision, but rather how its impact compares to other potential similar programmes.

The DFN leadership expressed concerns about these challenges and the robustness of the design, especially highlighting the challenge of comparing DFNPS to other provision, and the efforts which would be required to develop new relationship with comparator colleges. Overall, like for other methods presented in this section, with more robust design alternatives, we cannot recommend that the evaluation proceeds following this approach.

Appendix G: Options for outcome data access

The primary outcome for a full impact evaluation would be a measure of paid employment. There are a number of sources from which this data can be captured for a full impact evaluation. The table outlines these, including some of the advantages and disadvantages with each of those sources.

Options for sourcing primary outcome¹¹

Source	Advantages	Disadvantages
From DFNPS	The evaluator would only require a Data Sharing Agreement (DSA) with DFNPS, and not with all colleges (see option below), which would reduce costs.	DFNPS would have to amend their DSA with sites and hold personal details from participants. They are currently averse to holding this data. It would also require DFNPS to expand their data collection to cover control/comparator participants.
From colleges	The evaluator would have control over compilation of the data, and data could be collected after the internship ends, enabling faster reporting. It would not require resources from DFNPS.	This is resource intensive for the evaluator and colleges, given the amount of DSAs that might be needed. The data quality and granularity can be inconsistent across colleges. Since this data is not required to be reported to the ILR as of 2024/25, quality and consistency of data may vary greatly by provider, particularly for control/comparator participants.

¹¹ Note that we previously considered accessing employment outcome via the Individualised Learner Record (ILR), which has fewer barriers to access than LEO; however, it is our understanding that providers are no longer required to report this information to the ILR as of 2024/25.

Via the Longitudinal Educational Outcomes (LEO) dataset	There is no need to engage with colleges. LEO provides rich employment data, that enables tracking beyond the internship period. It tracks sustained employment outcomes (incl. earnings) longitudinally.	There is a significant time lag which would affect the evaluators' access to relevant data on time. Access to LEO for evaluation purposes (specifically the ingestion of a research participant identifier) is still being tested.
---	---	--

Given DFNPS's current reluctance to hold participants' personal identifiers and the lengthy timelines and access complexities associated with LEO data, the most feasible approach is to collect primary outcome data directly from colleges. This approach allows for timely collection of employment outcomes that align with DFNPS's definition of "paid and sustainable" work. This approach would also enable faster reporting than would be possible with administrative datasets. Depending on their existing data collection practices, we may need to ask colleges to expand their data collection to participants in the control/comparator group; if control/comparators were sourced from colleges with no DFNPS involvement, we would need to assess their current employment outcome data collection.

Softer outcomes, such as improved wellbeing, self-confidence, or social and communication skills, would have to be collected via surveys with young people after the time-point where supported internships were completed, or qualitatively via an Implementation and Process Evaluation (IPE).

Appendix H: Economic analysis

We have conducted an Economic Analysis to provide Youth Futures with a clearer picture on the Value for Money (VfM) associated with the evaluation of DFN Project SEARCH (DFNPS). Our aim is to calculate Minimum Meaningful Effect Sizes (MMES) which we define as the ones we expect to yield benefits at least equal to the costs associated to the internships, assuming the programme has a positive effect on employment (the primary outcome).

Costs associated to DFNPS

To calculate the unit cost of the internship, we consider two scenarios as suggested by Youth Futures¹²:

- £12,000 (proposed conservative lower bound)
- £14,000 (proposed conservative upper bound)

Benefits associated to DFNPS

If the supported internship causes a positive effect on employment, we focus on the Benefits associated to a young person avoiding being Not in Education, Employment or Training (NEET).

The GMCA Unit Cost database (2022/23) reports the following benefits per individual:

- **Economic benefit**, which refers to the broader value to society (including productivity). It is estimated at an annual savings of £11,474 per person for the economic cost associated with NEET.
- **Fiscal benefit**, which refers to the direct impact on the public purse (benefits, services, tax). It is estimated at an annual savings of £5,520 per person for the fiscal cost associated with NEET. This cost includes £92 for the Personal Independence Payment (PIP).

Bringing those estimations to current prices, we estimate:

- Benefit per employed-year (Economic): £12,721

¹² We also produced estimations for cost scenarios of £7 500 and £9 500 however it was decided that the other scenario is a better representation the resources required for the programme.

- Benefit per employed-year (Fiscal): £6,120

These was calculated inflating the 2022/23 estimations by 3.5% per year for 3 years.

Modelling

We produced simulations for return of investment (ROI) considering persistence of employment considering 3-year and 5-year scenarios. Persistence of employment measures how likely someone is to remain in the workforce over a specific duration after they've initially entered it.

We used two approaches to persistence:

- An asymptotic decay in the probability of remaining in employment, with half of those remaining in employment dropping out at each year. The specific profile of persistence is 100%, 50%, 25%, 12.5%, 6.25%.
- A diminishing decay, where half of the previous loss is experienced at each time point. The specific diminishing profile of persistence is 100%, 60%, 40%, 30%, 25%.

Based on this information, we simulated the effect sizes required to have a break-even impact of the programme (ROI=0%). For effect sizes, we focus on the percentage point difference in the likelihood of employment, that equals the costs and benefits (both economic and fiscal) associated with the supported Internships. This assumes a control group with a probability of employment of 0%. Our scenarios are included in *Table 17*, below.

Table 17: Estimated impacts for equal Cost and Benefits associated to the programme (percentage point difference in employment).

Cost per participant	Persistence profile	Effective employed-years per participant	Break-even Impact size (Economic)	Break-even Impact size (Fiscal)
£12,000	3y - Asymptotic	1.75	0.54	1.12
£12,000	3y - Diminishing	2.00	0.47	0.98
£12,000	5y - Asymptotic	1.94	0.49	1.01

£12,000	5y - Diminishing	2.55	0.37	0.77
£14,000	3y - Asymptotic	1.75	0.63	1.31
£14,000	3y - Diminishing	2.00	0.55	1.14
£14,000	5y - Asymptotic	1.94	0.57	1.18
£14,000	5y - Diminishing	2.50	0.44	0.92

To explain our estimations, we produced the following example:

Example back-of-the-envelope calculation of £12,000 cost per participant with a 3-year persistence under an Economic perspective:

Assuming the participant has a probability of employment of 100% (equivalent to 100% of the treatment participants 0% of the control participants getting into employment).

- Benefit per employed year (in 2025/26 prices) = (£11,474*(1+0.035)^3)) or £12,721
- Employed-years per participant (3-year profile using asymptotic persistence):
Total employed years = (year1) 1 + (year2) 0.5 + (year3) 0.25 = 1.75
- Benefit = £12,721*1.75 = £22,261.75
- **Cost/Benefit ratio:** £12,000/£22,261.75 = 0.539

Given a Cost Benefit ratio of 0.54, for cost and benefits to break even, we would expect the probability of being employment to be 54 percentage points. In the context of an RCT, this correspond to a scenario where there are 100 individuals in the control group, and none gets a job, and from the 100 individuals in the treatment group, 54 get a job.

54 pp will be therefore the MMES, as lower impacts than this, will represent an economic loss associated to investing in the programme, in the example. For reference, the effect associated with DFNPS in the US trial was 56 pp.

Appendix I: Supplementary Statistical Analysis

Table 18: Regression analyses exploring internship offers and acceptance

	Model 1: Rubric Score differences for those with and without offers (OLS)	Model 2: Predictors of Internship offers (Logit)	Model 3: Predictors of Internship acceptance (Logit)
Internship offered	10.989***	-	-
	(2.216)	-	-
Rubric score	-	0.097***	0.059*
	-	(0.023)	(0.025)
Age	-0.247	-0.001	-0.149
	(0.441)	(0.132)	(0.128)
Previous employment	8.261***	-0.524	-0.414
	(1.815)	(0.606)	(0.601)
Confidence score	1.512	-1.777*	-0.052
	(2.555)	(0.862)	(0.740)
Social skills score	-1.852	1.318	1.592
	(3.429)	(1.018)	(1.006)
Health & wellbeing	-0.632	0.299	-1.340

	Model 1: Rubric Score differences for those with and without offers (OLS)	Model 2: Predictors of Internship offers (Logit)	Model 3: Predictors of Internship acceptance (Logit)
	(3.213)	(0.967)	(1.038)
Overall health score	-0.010	0.007	0.009
	(0.039)	(0.012)	(0.011)
Financial independence score	-1.390	-0.993	-0.072
	(1.842)	(0.620)	(0.561)
Constant	83.080***	-3.803	-1.556
	(12.683)	(4.148)	(4.264)
Num.Obs.	180	180	143
R2	0.235		
R2 Adj.	0.189		
Dependent variable:	Rubric score	Offered internship	Accepted internship
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$			
Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.			
For modelling purposes, respondents with outstanding (i.e., pending or unconfirmed) offer information were coded as not offered.			

Table 19: McCrary density test —with MSE bandwidth and robustness check

Statistic	Default bandwidth	Bandwidth ± 10
Cutoff value	75	75
N (full sample)	401	401
N (left of cutoff)	130	130
N (right of cutoff)	271	271
Effective N (left)	119	66
Effective N (right)	262	146
Bandwidth (left)	22	10
Bandwidth (right)	22	10
T-statistic (jackknife)	3.239	0.866
P-value (jackknife)	0.001	0.386
Tests for manipulation of the running variable (rubric score) around the cutoff of 75. Jackknife standard errors used. Default bandwidth selected by MSE-optimal procedure.		

Table 20: Imbens–Kalyanaraman (IK) bandwidth output

Statistic	Value
IK MSE-optimal bandwidth	36.22
Suggested cut-off	75.00
Lower bound of bandwidth window	38.78
Upper bound of bandwidth window	111.22
Observed minimum of rubric score	37.00
Observed maximum of rubric score	100.00

Statistic	Value
Total observations (non-missing)	377.00
Observations within bandwidth	376.00
Observations below cut-off (within bandwidth)	116.00
Observations above cut-off (within bandwidth)	260.00
Observations outside bandwidth	1.00
Share of sample within bandwidth (%)	99.70
Source: authors' calculations based on rubric score data	