



Department
for Work &
Pensions



Evaluation of the Innovation Fund pilot

Quantitative assessment of impact and social
return on investment

September 2018

Research Report 956

A report of research carried out by National Centre for Social Research on behalf of the Department for Work and Pensions

Crown copyright 2018.

You may re-use this information (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence. To view this licence, visit <http://www.nationalarchives.gov.uk/doc/open-government-licence/> or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or email: psi@nationalarchives.gsi.gov.uk.

This document/publication is also available on our website at: <https://www.gov.uk/government/organisations/department-for-work-pensions/about/research#research-publications>

If you would like to know more about DWP research, please email: Socialresearch@dwp.gsi.gov.uk

First published September 2018.

ISBN 978 1 5286 0355 3

Views expressed in this report are not necessarily those of the Department for Work and Pensions or any other Government Department.

1. Acknowledgements

This report sits within the wider evaluation of the Innovation Fund (IF) pilot, commissioned by the Department for Work and Pensions (DWP). The evaluation was carried out by a number of organisations and individuals, including Insite Research, NatCen Social Research and Prof. Stephen McKay (University of Lincoln).

The authors gratefully acknowledge the support of colleagues at the DWP, in particular Yvonne Smith, Francesco Arzilli and Alex Snowsill. This report also benefited from the advice of a Steering Group, composed of Mike Daly, Agnieszka Nelson and Andrea Lee (DWP), Prof. Stephen Morris (Manchester Metropolitan University), James Magowan (Cabinet Office), and Eleanor Carter and Mara Airoidi (GO Lab, University of Oxford).

Thanks are also due to Emily Tanner (NatCen), for the support provided throughout this project, and to a number of people who previously worked on this evaluation when they were employed at NatCen, in particular Sarah Tipping, Francisco Moraiz, Hash Ahmed, Rossella Icardi and Nico Jabin. The report benefited from helpful discussions with Prof. Jeffrey Smith (University of Michigan), Karl Ashworth (NatCen) and Rita Griffiths (Insite Research).

2. The Authors

Sergio Salis is a Research Director at NatCen Social Research. He has evaluated the impact of a wide range of public programmes and policies, including the DWP's Pathways-to-Work programme, using quasi-experimental research methods.

Robert Wishart is a Researcher at NatCen Social Research. He is a trained economist and has previously conducted research on EU labour market trends, under-employment and the gender pay gap.

Stephen McKay is a Distinguished Professor in Social Research at the University of Lincoln and Associate Fellow of the Third Sector Research Centre. He has conducted cost-benefit analyses and analysis of social return on investment for a number of clients, including for vinspired.

Executive summary

Background

This report presents the findings from the quantitative evaluation of the Innovation Fund (IF) pilot, a £30million programme delivered between April 2012 and November 2015 to support young people aged 14 or over who were considered disadvantaged or at risk of disadvantage. The IF pilot was comprised of ten projects, which were commissioned in two rounds (Round One started in April 2012 while Round Two commenced operation in November 2012) and used a Social Impact Bond (SIB) model. SIBs are contracts for social outcomes, in which a delivery organisation receives funding from a social investor to deliver services, and government pays for the outcomes which are achieved as a result. The pilot aimed to help participants to re-engage with education, training and employment, reducing their longer-term dependency on benefits. Additionally, it aimed to build capacity within the social investment market, testing how to effectively develop and implement SIBs.

This report concludes a wider evaluation which included the qualitative assessment of the early implementation of the IF (Thomas and Griffiths, 2014) and of its delivery in the final year of operation (Thomas *et al.*, 2016).

Methodology

The quantitative evaluation of the IF pilot included an impact assessment and a Social Return on Investment (SROI) analysis. The impact assessment consisted of two separate analyses: one estimated the IF impact using mostly information from a survey of IF participants and a matched sample of non-participants, while the other relied exclusively on administrative data from the National Pupil Database (NPD) and Individualised Learner Record (ILR). The SROI analysis relied on impact estimates and financial information (invoicing and contractual data) supplied by DWP.

The survey-based analysis explored the impact of the pilot on the likelihood of Round One participants aged 14-18 being in education, employment or training around one year after starting on the pilot. The analysis based on administrative data investigated the IF impact on the likelihood of participants from either rounds aged 14-15 experiencing three educational outcomes (first NQF level 1, 2 and 3 qualifications) and two behavioural outcomes (persistent absenteeism and exclusion from school). The analysis of administrative data estimated impacts separately for three cohorts of participants (starters in November 2012-April 2013, November 2013-April 2014 and November 2014-April 2015), and impacts were explored for up to three years after individuals started on the programme (with follow-up being longest for the earliest cohort).

Both impact analyses used a Propensity Score Matching (PSM) approach to identify a comparison group, that is, a subset of non-participants (young people from areas not covered by the IF provision) similar to IF participants in several important respects so as to provide an indication of how participants would have fared had they not taken part in the IF pilot. (This method relies on observable data which are assumed to fully capture the characteristics of young people who were recruited to participate in the scheme and on comparable information being available for young people who did not receive the service.) The impact of the programme was then

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

estimated as the difference between the proportion of participants experiencing a specific outcome and the corresponding proportion observed among those in the comparison group.

The SROI analysis estimated the benefits produced by the pilot, and explored the extent to which these exceeded programme costs. The costs were based on funds actually paid by DWP, while the benefits reflected DWP's Willingness to Pay (WTP) for particular outcomes which resulted in subsequent expenditure savings.

Key findings

Due to data limitations, not all personal characteristics and circumstances associated with a greater probability of being or becoming NEET could be observed. Consequently, it is possible that the PSM approach failed to produce a comparison group fully comparable to programme participants. This in turn implies that the following findings from both the survey-based analysis and the analysis of administrative data should be treated with some caution.

The findings from the **survey-based impact analysis** can be summarised as follows:

- Around one year after starting the IF pilot, the proportion of participants who were in school or college was lower than the comparator group, likewise the proportion in a paid job. This finding suggests a negative impact of the IF pilot in that more young people would have been in education or employment had they not participated in the programme. On the contrary, the IF was found to have raised the proportion of participants in training, suggesting a positive impact.
- Young people could have been involved in multiple activities at the time of the survey (e.g., they could have been in employment but also doing a traineeship). When a more aggregated outcome measure was used, the results indicated that the IF pilot reduced the likelihood of participants being in employment, education or training (EET) one year after starting the programme. This negative impact was greater for older IF participants (those aged 16 or over) than younger ones (14 and 15 year olds).

The **impact analysis based on administrative** data found that:

- The IF pilot helped participants achieve a first NQF level 1 qualification, mostly one year after they started on the programme: the proportion of participants achieving this qualification was found to have increased as a result of the programme. This is consistent with findings from the IF qualitative evaluation (Thomas *et al.*, 2016).
- However, the programme also reduced the proportion of participants from all cohorts who achieved their first NQF level 2 and level 3 qualifications (for each cohort, the largest reduction was observed in the latest follow-up period). This indicates a negative impact in that IF participants were less likely to achieve these qualifications as a result of taking part in the pilot.
- Similar to the survey-based analysis, estimates of the impact analysis should be interpreted with caution because they could arise, at least in part, from methodological limitations arising from data constraints. However, the negative impacts could also reflect the incapacity of some projects to deliver staged outcomes, as suggested by Thomas *et al.* (2016). Cash flow requirements

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

might have led some projects to 'trade' the most ambitious individual outcomes (achievement of higher-level qualifications) with softer ones (improvement in attendance and attitude towards school).

- Positive impacts on NQF level 1 qualifications were higher and negative impacts on NQF level 2 qualifications more contained for later compared to earlier cohorts. This could possibly reflect a learning effect whereby programme implementation improved over time.
- Caution should be exercised when interpreting the estimated impact of the IF pilot on the likelihood of participants being persistently absent or excluded from school (negative in most cases). In addition to the possible limitations of the PSM approach discussed above, the measures used to capture behavioural outcomes (persistent absenteeism/exclusion from school in the academic year following the year in which pupils started on the IF pilot) were more stringent than those used to define the outcomes paid for by the programme (improvement in behaviour/persistent truancy demonstrated for a minimum of 13 continuous weeks). Consequently, they are likely to have largely underestimated the benefits of the pilot.

The findings from the **SROI analysis** are as follows:

- Using the WTP metric (which was based on direct cashable savings) to measure the benefits of the pilot, the programme achieved value for money: the ratio of benefits to cost was around 1.3 for Round One projects and 1.25 for Round Two projects.
- However, these ratios are much smaller than those produced for similar programmes, and lower than the kinds of ratios that would be generated using wider social benefits (notably those derived from the 'well-being valuation' approach). Using estimates based on these wider numbers would, before looking at the 'deadweight' within the programme, have generated SROI ratios 4-5 times higher. Data on less tangible outcomes would push this number still higher.
- The findings from the impact evaluation imply that many outcomes would have been achieved without the programme, which suggests that the real SROI ratios would be below 1. Taking into account the wider social impacts increases the total size of the financial effects of the policy, but since this cannot be specifically attributed to the programme, adopting a broader perspective does not produce a more positive SROI ratio.

The final conclusions need to account for the limitations of the impact estimates, and should therefore allow for some degree of uncertainty.

List of abbreviations

DWP	Department for Work and Pensions
IF	Innovation Fund
ILR	Individualised Learner Record
MI	Management Information
NEET	Not in Education, Employment or Training
NPD	National Pupil Database
NQF	National Qualifications Framework
PbR	Payment by Results
PSM	Propensity Score Matching
QCF	Qualifications and Credit Framework
SIB	Social Impact Bond
SROI	Social Return on Investment
WTP	Willingness to Pay

Contents

- Acknowledgements 3
- The Authors 4
- Executive summary 5
- List of abbreviations..... 8
- 1 Introduction..... 12
 - 1.1 Background 12
 - 1.2 The pilot projects..... 12
 - 1.3 The social investment model..... 13
 - 1.4 Aims of the report..... 14
 - 1.4.1 Impact assessment 14
 - 1.4.2 Social Return on Investment analysis 15
 - 1.5 The quantitative methodology 15
 - 1.6 Report outline..... 16
- 2 Quantitative methodology..... 17
 - 2.1 Overview of the data used 17
 - 2.1.1 Survey data 17
 - 2.1.2 Management Information (MI)..... 18
 - 2.1.3 Administrative data..... 18
 - 2.1.4 Cost data 19
 - 2.2 Sample identification 19
 - 2.2.1 Impact analysis based on survey data 19
 - 2.2.2 Impact analysis based on administrative data 21
 - 2.3 Timing of the IF pilots and the survey 23
 - 2.4 Impact estimation 24
 - 2.5 The PSM approach to identifying the comparison group 25
 - 2.6 Application of the PSM to the impact analysis 26
 - 2.7 SROI analysis 26
 - 2.8 Data considerations 26

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

- 3 Results from the impact analysis based on survey data 28
 - 3.1 Introduction 28
 - 3.2 Impact results 28
 - 3.2.1 Impact on current activities 29
 - 3.2.2 Impact on the EET status 30
 - 3.2.3 Impact by age of participants at IF start 31
 - 3.3 Discussion of findings 32
- 4 Results from administrative data: Overall impact 34
 - 4.1 Introduction 34
 - 4.2 Impact on educational qualifications 35
 - 4.2.1 Impact on the likelihood of achieving a first NQF level 1 qualification 35
 - 4.2.2 Impact on the likelihood of achieving a first NQF level 2 qualification 37
 - 4.2.3 Impact on the likelihood of achieving a first NQF level 3 qualification 39
 - 4.2.4 Comparing impacts across cohorts 40
 - 4.3 Impact on school attendance and behaviour 41
 - 4.4 Further considerations 43
- 5 Results from administrative data: Subgroup impacts 44
 - 5.1 Introduction 44
 - 5.2 By-round impact 44
 - 5.2.1 Impact on the likelihood of achieving a first NQF level 1 qualification 45
 - 5.2.2 Impact on the likelihood of achieving a first NQF level 2 qualification 46
 - 5.2.3 Impact on the likelihood of achieving a first NQF level 3 qualification 48
 - 5.2.4 Impact on school attendance and behaviour 49
 - 5.3 By-gender impact 50
 - 5.3.1 Impact on the likelihood of achieving a first NQF level 1 qualification 50
 - 5.3.2 Impact on the likelihood of achieving a first NQF level 2 qualification 51
 - 5.3.3 Impact on the likelihood of achieving a first NQF level 3 qualification 53
 - 5.3.4 Impact on school attendance and behaviour 54
 - 5.4 Discussion of the findings and further considerations 55
- 6 SROI analysis 57
 - 6.1 Background 57

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

- 6.1.1 Social Return on Investment (SROI)..... 57
- 6.1.2 Social Impact Bonds (SIBs) 57
- 6.2 Costs and benefits of the IF 59
 - 6.2.1 Direct costs and cashable benefits 59
 - 6.2.2 Provider-level analysis 63
 - 6.2.3 Other approaches to monetising the IF benefit..... 64
- 6.3 Calculating ‘additionality’ 65
 - 6.3.1 Deadweight and the impact estimates 65
 - 6.3.2 Further caveats 66
- 7 Conclusions..... 67
 - 7.1 Impact evaluation 67
 - 7.2 SROI analysis 69
 - 7.3 Overall conclusions 69
- References 70
- Appendices 72

1 Introduction

1.1 Background

The Innovation Fund (IF) pilot initiative was announced in 2011 as part of a package of measures aimed at reducing youth unemployment. A fund of approximately £30 million, it was set up to support ten social investment projects based on a Payment by Results (PbR) model, which stipulates that the investors (or intermediaries) are paid exclusively on the basis of the achievement of pre-agreed outcomes.

The IF pilot had three key objectives:

- Supporting young people aged between 14 and 24 who were either disadvantaged or at risk of disadvantage, helping them participate in education, training and employment, and in so doing reducing their longer-term dependency on benefits;
- Testing the ability of social investment models to produce benefit savings and/or other wider fiscal and social benefits, and deliver Social Return on Investment (SROI); and
- Helping the development of the social investment market and smaller delivery organisations, and producing credible evidence that can guide similar social investment funded initiatives.

1.2 The pilot projects

The Department for Work and Pensions (DWP) awarded ten IF pilot contracts over two rounds. The first round covered six projects which were awarded in spring 2012 and became operational in April that year. The four projects included in the second round were awarded in late summer 2012 and started operating from November 2012. The projects ended around three years after their start date (Round One projects in April 2015 and Round Two projects in November 2015), and outcomes were monitored (and paid) for a further six months.

The pilot projects supported a broad range of vulnerable and disadvantaged young people characterised by a diverse range of problems, including truancy and disengagement from school, learning difficulties, and poor numeracy and literacy skills. While young people not in education, employment or training (NEET) and those at greatest risk of becoming NEET were a main target of the IF pilot initiative (the programme was designed to be a corrective intervention for the former and a preventative measure for the latter), the programme also supported other vulnerable young people. These were young offenders, drug users, gang members, young parents, care leavers and those living with a mental health condition.

The ten pilot projects were spread out across different geographical areas of the UK and displayed diversity in terms of the number of delivery bodies, type of investors, participants' age, number of young people supported (size), and groups targeted.¹

¹Pilot projects in Round One areas targeted young people aged between 14 and 24, the exact age coverage depending on the project, while Round Two projects worked exclusively with young people aged 14 and 15. See Appendix B 'The 10 pilot projects at a glance' of the qualitative reports (Thomas and Griffiths, 2014; and Thomas et al., 2016) for a more detailed description of each of these projects.

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

For the purpose of this report, the project names are those of the organisations which entered into a contractual arrangement with the DWP for the delivery of the programme (also known as Special Purpose Vehicles). Table 1 below lists all of the projects and the geographical areas covered.

Table 1.1 Projects and associated geographical areas

Projects	Area covered
Round One	
Nottinghamshire City Council	Nottingham
Advanced Personnel Management (APM) Ltd	Birmingham and Solihull
Triodos	Greater Merseyside
Indigo Project Solutions	Perth and Kinross (Scotland)
Links to Success	Stratford, West Ham, Canning Town, Custom House and Royal Docks in the borough of Newham, and Cathall and Cann Hall in the borough of Waltham Forest
Private Equity Foundation (PEF)	Shoreditch, East London, including Hackney, Islington and Tower Hamlets
Round Two	
Teens & Toddlers	Greater Manchester, including Bolton, Manchester, Oldham, Salford and Tameside
3SC	Cardiff and Newport (Wales)
Energise Innovation Ltd	Thames Valley
Prevista	Six boroughs in West London: Brent, Ealing, Hammersmith and Fulham, Hounslow, Westminster and Haringey

This evaluation relied on National Pupil Database (NPD) data to assess the impact of the IF pilot. This data source contained a wide range of variables needed to implement the quantitative approach to estimate impacts but covered only pupils in England. This is the reason that projects delivered in Scotland (Indigo Project Solutions) and Wales (3SC) were excluded from the analysis.²

1.3 The social investment model

The social investment model, on which the IF pilot was based, aims to produce both social and financial returns. However, it is the social returns that constitute the primary objective of this model: while any financial return is generally considered as desirable, social investment models are specifically designed to improve outcomes for a particular social group. In the specific case of the IF pilots, the social group targeted included disadvantaged young people who were (or at risk of becoming) NEET, and the outcomes to be improved encompassed jobs and individual educational achievements and behaviours reflecting improved chances of future employment.

²The IF survey covered only English projects. For comparability reasons, it was decided that the impact analysis based on administrative data would also be limited to England.

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

The pilot projects were funded through a Social Impact Bonds (SIB) mechanism. A SIB is a contract between an investor and the public sector, where the former provides the monetary resources needed to implement the socially-relevant project and the latter (who commissions the contract) commits to re-pay them for improved social outcomes which result in public sector savings. A defining feature of SIBs is that repayment is contingent upon pre-agreed social outcomes being achieved (SIBs are included within the broad category of PbR models).

The outcomes paid to the investors by the DWP under the IF pilot included young people's entry to, and sustained maintenance of, first employment, and indicators of improved future employability chances. These indicators included improved attitude, behaviour and attendance at school (through stopping persistent truancy and poor behaviour at school), and the achievement of certain educational qualifications (notably, first Qualifications and Credit Framework (QCF) Entry-level and NQF levels 1 to 4).

1.4 Aims of the report

This report sits within the wider evaluation of the IF pilot commissioned by the DWP, and follows a qualitative evaluation strand of research on the early implementation of the programme (Thomas and Griffiths, 2014) and its delivery in the final year of operation (Thomas *et al.*, 2016). Its chief aim is to illustrate the findings from the quantitative evaluation of the IF pilot, which included two main components: an impact assessment and an SROI analysis.

1.4.1 Impact assessment

The impact assessment comprised two analyses: one used mainly survey data and the other relied exclusively on administrative data. Both analyses sought to gauge evidence as to whether the programme has been effective in achieving relevant participants' outcomes related to employment and increased future chances of employment.

Survey-based impact analysis - a key objective of this evaluation was to provide a measure of the impact of the IF pilot on relevant outcomes for participants both at an early stage of the programme and at the end of the three and a half year implementation period. The information gathered from the survey of IF participants and non-participants made it possible to assess the impact of the pilot on the likelihood of young people being in employment, education or training around one year after they started participating in the pilot. Unfortunately, the second wave of the survey achieved sample sizes which were too small to provide reliable findings and therefore exploring longer-term outcomes was not possible using information from respondents.

The survey-based impact analysis explored the impact of the IF pilot for young people aged between 14 and 18 who participated in any of the Round One projects delivered in England.

Administrative-based impact analysis - using administrative data sources meant that the impact of the IF pilot on programme participants was explored for educational attainments (achievements of a first NQF level 1, 2 or 3 qualification at different time points over a time span of up to three years following young people's start on the pilot) and behavioural outcomes (whether participants were persistently absent and excluded from school over the year after participants started on the pilot projects).

Beyond providing a measure of impact for the IF pilot programme as a whole (the 'overall' impact), impacts were also assessed by rounds of commissioning and by gender of the participants.

The impact analysis which used administrative data explored the effect of the IF pilot on youngest participants (those aged 14 or 15) from English projects from both rounds. The focus on the youngest participants reflected the reality of the programme delivery (several Round One projects shifted towards this age group because, as these people were still at school, they were easier to access and recruit than older individuals), and also enhanced comparability of impacts between the two rounds of commissioning (Round Two projects supported exclusively 14 and 15 year olds).³ The interest in younger participants also meant that the evaluation of the IF pilot which used administrative data could not include the Nottingham City Council project, which only supported young people aged 16 or more.

Whereas the impact estimates produced by means of the analyses based on survey and administrative data (presented separately in the report) are not always directly comparable, some common findings emerged. These potentially provide the basis for additional research on the IF pilot and future investigation of the effectiveness of similar interventions, and contribute to the debate about the effectiveness of programmes funded by means of SIBs.

1.4.2 Social Return on Investment (SROI) analysis

The SROI analysis aimed to estimate the range of benefits produced by the IF pilot, including any less tangible benefits, and to consider how far these exceeded the programme's various costs. The financial information, relating to costs and to the number and types of outcomes achieved, was supplied by the DWP. The SROI also drew on the results from the impact assessment to consider how far the apparent benefits of the programme can be attributed to the actions of the providers, rather than being outcomes that might have happened even in the absence of the programme.

1.5 The quantitative methodology

The impact of the IF pilot on programme participants was estimated as the difference between the average outcome observed among participants and the average outcome observed among a similar group of non-participants (the 'comparison' group). The latter outcome provided a measure of the 'counterfactual,' that is, the (hypothetical) outcome that participants would have experienced had they not started the programme. The comparison group was obtained by implementing the Propensity Score Matching (PSM) methodology, which entailed pairing (or 'matching') each IF participant with the most similar non-participant young person. Similarity was assessed across a range of personal characteristics captured by means of administrative and survey data.

³The qualitative impact evaluation (Thomas *et al.*, 2016) found that, after realising that the originally expected numbers of outcome targets per participant would not be achieved, many projects recruited additional schools and refocused their support onto young people in compulsory education to increase the throughput of recruits. As the authors reported 'Several projects whose original bids focused strongly on recruiting young people who had left school and were already NEET increased the number of starts by working more with schools and with younger school-aged cohorts (Thomas *et al.*, 2016, page 40).'

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

Two sets of impact estimates were produced: one based solely on administrative data and the other using predominantly survey data. In the former, both outcome and control variables (i.e., variables used to control for similarity between participants and non-participant young people in the PSM) were sourced from the NPD and Individualised Learning Records (ILR). The latter relied on survey data to provide outcome and control variables (however, some control variables from the NPD were also used). Survey estimates had the advantage of covering a wider range of outcomes (e.g., in addition to educational outcomes, employment- and training-related outcomes were explored) and using a richer set of control variables to implement PSM. On the other hand, estimates based on administrative data benefited from larger sample sizes and made it possible to investigate sub-group impacts (by round of commissioning and gender of IF participants).

The Social Return on Investment model seeks to estimate the monetised benefits of a given intervention, and to compare those benefits against the costs of the programme. It has similarities with cost-benefit analysis. In attributing benefits to the intervention, the impact study is used to consider the issue of deadweight (this concerns participants who experienced certain outcomes but would have achieved these outcomes anyway in the absence of the programme).

1.6 Report outline

This chapter set out the background of the IF pilot and introduced the quantitative evaluation which produced the findings presented in this report. The remainder of the report is organised as follows. Chapter 2 illustrates the data used to implement the impact assessments (analyses based on survey and administrative data) and the SROI analysis, and outlines the analytical methodologies underpinning them. Chapters 3 and 4 present the results of the impact analyses which used survey and administrative data, respectively. Chapter 5 shows the findings from the subgroup impact analyses (by round of commissioning and gender of programme participant) based on administrative data. Chapter 6 shows the findings from the SROI analysis. Finally, Chapter 7 concludes the report.

2 Quantitative methodology

This chapter describes the data used to estimate the impact of the Innovation Fund (IF) pilot, and gives an overview of how the samples of IF participants and non-participants used for the analyses based on survey and administrative data were obtained. In particular, the Propensity Score Matching (PSM) approach used to identify the comparison groups of non-participants, used to estimate impacts for both analyses, is illustrated. The analytical approach to assessing the Social Return on Investment (SROI) is also presented. Finally, data limitations potentially affecting the validity of the impact assessment and SROI analysis are reported.

2.1 Overview of the data used

The data used to assess the impact of the IF pilot came from different sources: a survey of IF participants and non-participants (implemented by NatCen), the IF's Management Information (MI) (collected by the programme contractors and provided by the DWP for analysis) and administrative data from the National Pupil Database (NPD) and Individualised Learner Record (ILR), which were obtained from the DfE. The SROI analysis used the results of the impact assessment (the impact estimates) and other data on costs. The data on costs, provided by the DWP, concerned financial transactions relating to the programme (i.e., the number of outcomes, the value claimed by the provider, and DWP's maximum Willingness to Pay (WTP) for particular outcomes). These data sources are described below.

2.1.1 Survey data

The impact of the IF pilot was assessed using data gathered from the first wave of a face-to-face survey of young people who participated in Round One of the pilot and a similar group of non-participant individuals. The non-participant sample was selected to resemble IF participants in some important respects, notably demographic- and education-related characteristics observed prior to starting the IF (captured through data sourced from the NPD) using the Propensity Score Matching (PSM) approach (details are given in section 2.5).

By surveying IF participants and non-participants, it was possible to collect further information about their characteristics, personal circumstances and activities prior to the pilot. These additional baseline characteristics included individuals' personal details (e.g., date of birth and whether born in the UK), and parent- and family-related characteristics (number of siblings and whether either of their parents stayed in education after the age of 16). Observing a wider range of personal characteristics

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

made it possible to further refine the surveyed group of non-participants by selecting, among all individuals in this group, those who matched more closely the profile of IF participants.⁴

Being able to compare IF participants to a refined, similar group of non-participants was key to building up a picture of how the IF had an impact on programme participants. The survey collected information on young people's economic activities subsequent to IF participation (whether they were working, going to school or attending a training course or apprenticeship) so that the impact of the pilot on these outcomes could be explored.

2.1.2 Management Information (MI)

The MI data provided to NatCen to implement the impact assessment contained the vast majority of young people who took part in the IF pilot projects as it covered all those who started either a Round One or Round Two project between April 2012 and June 2015. The MI provided information about all Round One starters and personal details for those who started the IF Round Two between July and November 2015, but excluded young people who started the pilot in the last five months of Round Two.

The MI provided a wide range of IF participants' personal details, including their forenames, surnames, dates of birth and addresses. Using these four pieces of information, individual administrative records from the NPD and ILR were linked back to programme participants. It also indicated the date at which each young person started the IF pilot and the delivery body responsible for the provision of the service.⁵ Among all young people recorded in the MI, only those who gave their consent to using their personal information to conduct statistical analysis were used in the impact assessment.

2.1.3 Administrative data

Administrative data were sourced from the NPD and ILR. The NPD includes comprehensive information about students who attend schools and colleges in England. For example, it contains variables indicating whether each young person had any Special Education Needs (SEN) and, if so, what action has been taken by schools or local authorities to address needs. From the NPD it was also possible to know whether pupils were eligible for Free School Meals (FSM). Information about individuals' demographic characteristics such as gender, ethnicity (whether white, black, Asian, mixed or from any other ethnic origin) and language (whether English was their first language) were sourced from this dataset. The NPD also includes pupils' test and exam results, prior attainments and progression at different stages for all schools in the state sector in England (Key Stages 2, 4 and 5 results were used for this report), and information about pupils' absenteeism and exclusions.

⁴Essentially, PSM was first implemented using data sourced from the NPD to identify the sample of non-participants to be surveyed together with IF participants; PSM was then re-applied using the original NPD variables and new personal information collected through the survey. The second round of matching was carried out to attempt to eliminate (or minimise) any residual difference between surveyed participants and non-participants affecting the outcomes of interest and participation in the IF pilot not captured through NPD data.

⁵The contractor under which each delivery body operated and the area covered was identified through a mapping document provided by the DWP. The Department also provided information about the round for each contractor/delivery body.

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

The ILR data contain information concerning young people aged 16 or more attending Further Education (FE) through publicly funded colleges, training organisations, local authorities or employers. They include records on qualifications being studied and the activities in which a young person is participating (e.g., work-based learning or apprenticeships). However, our analysis solely draws upon the qualification data, which are used to derive educational-related attainment variables (first NQF Levels 1, 2 and 3). Qualifications detailed in the ILR data tend to be predominantly vocational-based qualifications (such as NVQs), as opposed to more academic qualifications observed in the KS5 data files of the NPD.

2.1.4 Cost data

The data on costs were supplied by DWP. They consisted of financial records – invitation to tender, invoices – relating to the operation of the programme. The data may have been amalgamated at provider level, and identified the number of different outcomes achieved by each provider, and the amount paid by the DWP in return, based on the contracts between DWP and each provider. The willingness of DWP to pay for each outcome, which is also published data, was used to inform the analysis of benefits and costs within the SROI.

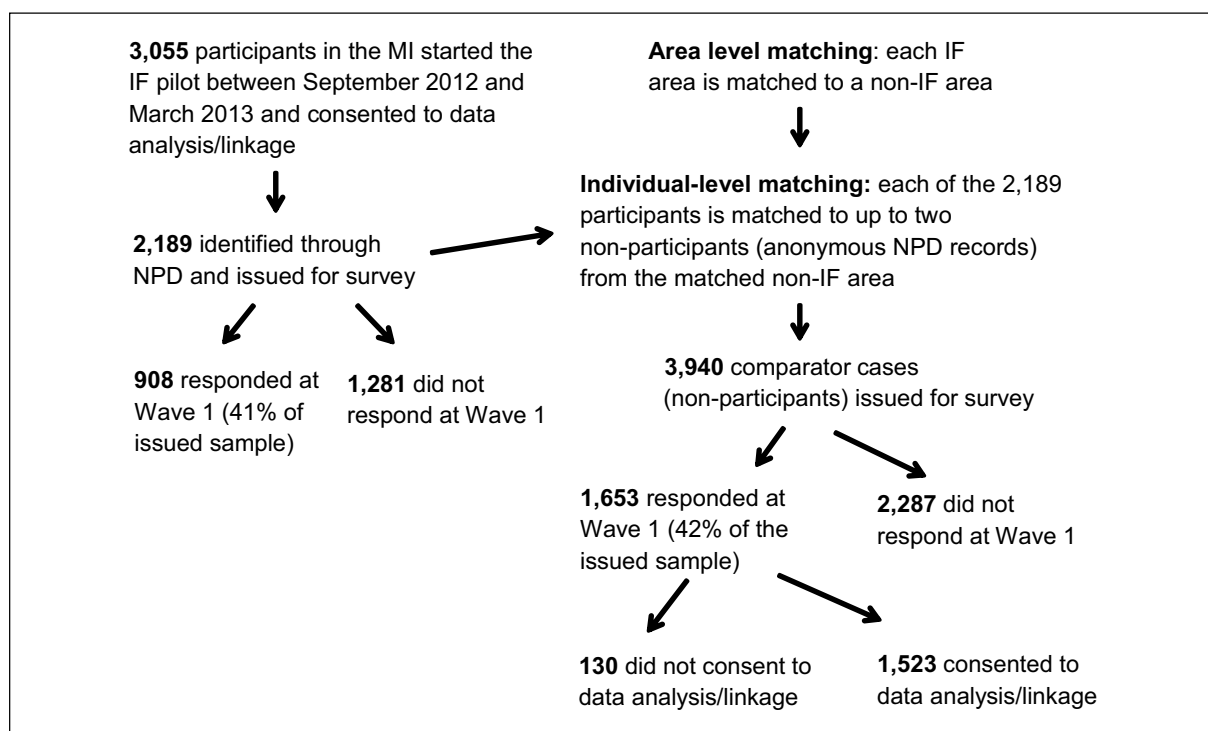
2.2 Sample identification

This section explains how the samples of participants and non-participants were constructed to implement the survey-based impact analysis and the impact assessment which used administrative data.

2.2.1 Impact analysis based on survey data

The Wave One IF survey collected detailed information from a sample of 908 IF (Round One) participants and a selected, matched sample of 1,523 non-participants. The flowchart in Figure 2.1 shows how these two samples were obtained.

Figure 2.1 Identification of the samples of survey participants and non-participants



The sample of 908 IF participants was identified using the pilots' MI. All young people recorded in the MI, as provided at the time the survey-based analysis was designed (3,055 cases), gave their consent to use their information for data analysis. Among the numerous pieces of information about young people who took part in the IF pilot projects, the MI contained programme participants' forenames, surnames, dates of birth and addresses, which were used to link individual NPD records to 2,189 (72 per cent) of the 3,055 programme participants. This is the size of the sample of IF participants issued for the survey. However, the sample size finally achieved was smaller than 2,189 as a survey response rate of 41 per cent for the participant group meant that only 908 IF participants could be used in the analysis.

The sample of 1,523 non-participants was obtained following a two-step PSM procedure. In the first step, among all the English LAs in which the programme had not been piloted (non-IF areas), those most similar to the pilot areas (or IF areas) were identified by matching (pairing) each IF area to a non-IF area based on similarity across a range of socio-economic characteristics of its population (e.g. level of deprivation and percentage of 14 to 16 years old young people who were NEET).⁶ In the second step, within each IF area every programme participant was matched (based on individual characteristics captured by NPD variables) to either one or two residents of the areas chosen as the comparator areas for that specific IF area. In this way, it was possible to identify a sample of 3,940 non-participant individuals to be surveyed alongside the 2,189 programme participants. The response rate to the survey for the non-participant sample was around 42 per cent (very similar to that achieved for the IF participant sample), and meant that

⁶Table A.1 in Appendix A shows the areas matched to IF areas.

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

of the 3,940 young people issued for the survey 1,653 responded. Among these 1,653 individuals, 130 did not consent to their survey data to be linked to NPD for analysis. Therefore, the final sample size available for the non-participant group was 1,523 cases.

The participant and non-participants samples were further reduced following a second round of individual-level matching (see section 2.6): 891 of the 908 respondents available from the participant sample were paired to a 'refined' sample of 891 of the 1,523 non-participants. The impact assessment based on survey data estimated impacts using outcome data for these 891 participants and 891 comparators.

2.2.2 Impact analysis based on administrative data

The sample of IF participants used for the impact evaluation based on administrative data was obtained linking NPD and ILR data to all programme participants recorded in the MI who gave their consent to data linkage and analysis (17,073 individuals).⁷ Personal identifiers (forename, surname, date of birth and postcode) for these individuals were sent to the DfE, who linked these cases to administrative records from the NPD and ILR. The fact that that 881 individuals were IF participants from Scottish projects and 680 from Welsh projects (not in scope for this evaluation), and 603 individuals had wrongly recorded (inexistent) postcodes, meant that the DfE received identifiers for 14,909 participants in English projects.

The matching rate was relatively high: 12,401 of these 14,909 individuals (83%) could be linked to (i.e., were found to have one or more records in) at least one among the various NPD and ILR data files. Data cleaning further reduced the size of the IF participant sample to 10,959 young people aged between 14 and 24 years at the time they started the IF pilot. The decision to focus on young people aged 14 and 15 at programme start (made after the data linking process was implemented by the DfE) meant that a subset of 5,597 individuals within this age band was initially considered for the impact analysis. However, the sample size actually used was smaller due to the analysis set-up (cohort approach) adopted. This entailed assessing impacts, separately, for three cohorts of IF participants (aged 14 or 15 at IF start):

- **2012/13 cohort:** started the IF between November 2012 and April 2013;
- **2013/14 cohort:** started the IF between November 2013 and April 2014; and
- **2014/15 cohort:** started the IF between November 2014 and April 2015.

All of the IF participants who started the pilot at a date which fell outside any of these time windows or, in spite of having a start date within these windows, were aged over 15 were excluded from the analysis.

Information about 879,176 young people resident in non-IF areas (unitary authorities defined according to the post-2009 local government restructuring) was also available to identify a comparison group of non-participants. Each of these individuals were previously assigned a pseudo-start date which were treated as equivalent to the start dates actually observed for participants, making it possible to compare the

⁷Note that these 17,073 participants started the IF pilot at some point between April 2012 and June 2015, and therefore include all 3,055 IF participants used to produce the survey-based impact estimates (individuals in this subset started the IF pilot between September 2012 and March 2013 - see Table 2.2).

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

outcomes of participants and non-participants to estimate the IF impact.⁸ Assigning pseudo-start dates to non-participants meant that three cohorts, defined in a similar way as for participants, could be identified also for this group. That is, the pseudo-start dates of the 879,176 non-participants fell within any of the periods among November 2012-April 2013, November 2013-April 2014 and November 2014-April 2015, and non-participants were aged 14 or 15 at IF pseudo start.

Table 2.1 shows the sample size of each cohort of IF participants and non-participants. In total, the three cohorts included 2,817 participants and 879,176 non-participants.

Table 2.1 Sample sizes (impact analysis based on administrative data)

	Participants	Non-participants
2012/13 cohort (November 2012-April 2013 starters)	1,258	286,419
2013/14 cohort (November 2013-April 2014 starters)	1,005	266,970
2014/15 cohort (November 2014-April 2015 starters)	554	325,787

Base: 2,817 participants and 879,176 non-participants.

Note that the very large size of the three non-participant cohorts made estimating impacts unfeasible (the calculation of the standard errors following implementation of the PSM required a very large amount of processing power). Consequently, it was decided to randomly select a 20 per cent sample from each of the three cohorts, and use them to estimate impacts.

There were specific reasons for choosing to include in the analysis all participants with an observed IF start date within the November-April window. Educational outcomes (NQF level 1, 2 and 3 qualifications) are normally achieved in the last term of each academic year (between May and July) but the exact date of award is not recorded in NPD data. As assessing the causal effect of the IF on these outcomes implies observing them after the start of the programme, it was decided to exclude all participants with a start date after April of the academic years in question as for these people it was not clear whether the outcome followed the IF pilot or the other way around.

The reason for selecting only participants with start dates from (and including) November was instead driven by the need to compare impacts between Round One projects (starting in April 2012) and Round Two projects (starting in November 2012). This also helped avoid possible time-related effects (Round One starters between April and October 2012 experiencing different impacts from starters in later months).

⁸The DfE provided anonymous NPD/ILR records for the entire population of non-participant individuals (residents in areas not exposed to the IF pilot) born between the 30th of June 1987 and the 10th of July 2001 (the earliest and latest date of birth observed among the IF participants). Each of these young people (almost six million cases) was randomly assigned a pseudo-start date (Appendix B provides details on how pseudo start dates were assigned to non-participants; see Orr et al. (2007) and Ainsworth and Marlow (2011) for previous applications of this methodology) and their age at (pseudo) IF start was calculated as the number of years which elapsed between their date of birth and pseudo start date. Those with a pseudo start date which fell within any of the time windows among November 2012-April 2013, November 2013-April 2014 and November 2014-April 2015, and aged either 14 or 15 at IF pseudo-start date (879,176 cases), were considered for analysis.

2.3 Timing of the IF pilots and the survey

As already mentioned in the first chapter of this report, the IF pilots were introduced in two rounds. Round One started in April 2012 and ended around three years after (in April 2015), while Round Two took place between November 2012 and November 2015. Whereas the impact evaluation which used administrative data used both participants in Round One and Round Two projects, the impact assessment based on survey data focused exclusively on Round One participants.

The summary distributions of the programme start dates for Round One participants and the subset of participants who responded to the survey (908 cases) is shown in Table 2.2. Survey respondents started the IF pilots in the time period between September 2012 and March 2013, and therefore this is the time span over which the distribution of start dates is also explored for the entire population of programme participants. In the MI, 3,055 cases were recorded as having started the pilot within this time period.

Table 2.2 Distribution of IF start dates

Date (month and year)	IF participants recorded in the MI (%)	IF participants who responded to the survey (%)
September 2012	19.9	27.4
October 2012	18.1	17.6
November 2012	11.8	10.2
December 2012	5.1	4.3
January 2013	15.4	15.2
February 2013	13.7	11.6

Base: 3,055 IF Round One participants observed in the MI and 908 IF Round One participants who responded to the survey.

As the table indicates, over one quarter (27.4 per cent) of the sample of programme participants who responded to the survey started the IF pilot in September 2012, and the proportion of participants progressively decreased over the next three months to reach its lowest point in December (when only 4.3 per cent of the IF participant sample started the programme). In the first three months of 2013, the monthly proportion of starters is between 11.6 and 15.2 per cent.

Comparing the distribution of start dates observed for the IF participants who responded to the survey to the correspondent distribution for all programme participants, we notice that to some extent the survey over-sampled IF participants who started in September 2012. The proportion of all participants who started the IF in this month is in fact lower (it being 19.9 per cent) than that observed for the IF participants who responded to the survey (27.4 per cent). This is mirrored by a slight under-sampling of programme participants in all months from October 2012 to March 2013.

Individuals in the participant sample were interviewed between the 23rd of October 2013 and the 23rd of January 2014. This means that, on average, around one year (363 days) elapsed between the date IF participants started a Round One project and

the date they completed the survey. The interviews for the sample of non-participants started at around the same time as participants (25th of October 2013) and ended four months later (17th of February 2014).

2.4 Impact estimation

This evaluation assessed the impact (or causal effect) of the IF pilot on a range of different outcomes for programme participants. The samples of programme participants and the outcomes used to estimate impacts varied depending on whether the impact analysis used survey or administrative data.

The analysis which used survey data estimated the impact of the IF pilot for a sample of 891 survey respondents who started any of the Round One English projects between September 2012 and March 2013, aged between 14 and 18. For these individuals, the analysis assessed the impact of the pilot on the likelihood of being involved in a range of activities related to employment, education and training at the time of the survey interview (which happened, on average, one year after the sample of participants started on the pilot).

The analysis based on administrative data examined the impact of the IF pilot for three cohorts of participants: young people who started an IF pilot project between November 2012 and April 2013, November 2013 and April 2014, and November 2014 and April 2015, being aged 14 or 15. For each of these cohorts, the analysis assessed the impact of the programme on the likelihood of achieving each educational qualification, among a first NQF level 1, 2 and 3, at different time points following participants' start on the pilot (the actual follow up period depended on the cohort explored). Additionally, the analysis explored the IF impact on the likelihood of participants being persistently absent and excluded from school, respectively, in the academic year after they started the pilot.

For each cohort of participants from the administrative data, and for the sample of participants who responded to the survey, the PSM approach (see next section) was employed to identify a 'comparison' group of non-participants.⁹ Each comparison group comprised a subset of individuals from the non-participant sample similar to participants with respect to a range of demographic and educational characteristics, and whose outcome were used to proxy the outcome that participants would have experienced had they not taken part in the IF pilot (this hypothetical outcome is known as the 'counterfactual').

The difference between the proportion of the participant cohort/sample who experienced a given outcome and the correspondent proportion for the comparison group (the counterfactual) provided an estimate of the impact of the IF pilot on the likelihood of participants experiencing this outcome.¹⁰ Essentially, the IF impact was calculated as the increase (or reduction, for negative impacts) in the proportion of participants who experienced the outcome of interest (e.g., achieving a specific qualification or being in employment) as a result of the pilot. This change was expressed in percentage points (ppts).

⁹PSM is a widely-used approach to estimating programme impacts. See, for example, Blundell and Costa Dias (2000) for technical details on this methodology.

¹⁰Using the terminology of the evaluation literature, an impact estimated in this way is known as the 'average effect of the treatment on the treated.'

2.5 The PSM approach to identifying the comparison group

The PSM approach is aimed at identifying a comparison group of non-participant individuals (in this case, young people from geographical areas other than those where the IF pilot was rolled out, and therefore unlikely to have been exposed to the programme) for whom the distribution of characteristics observed at the baseline (i.e., prior to participation in the pilot) is as similar as possible to the distribution for participants. The approach involved two steps. The first step entailed estimating, for each individual, the probability of participating in the IF pilot conditional on their baseline characteristics (this probability is called *propensity score*).¹¹ The characteristics considered for PSM (also called control variables or covariates) should include all determinants of participation in the programme which are also predictive of the outcomes of interest.¹² The second step involved pairing, or ‘matching,’ each participant to the non-participant with the closest propensity scores.¹³ The propensity score has the advantage of summarising all individual features in a single index, and therefore similar scores reflect similarity across the entire range of characteristics considered.

The PSM method aims to correct for selection bias: this is the error made when estimating impacts as simple differences between the outcomes observed among participants and (all) non-participants in the presence of relevant differences between the two groups. For example, we know that 43.2 per cent of the participant sample used for the impact analysis based on administrative data were eligible for Free School Meals while a much smaller proportion (15.7 per cent) was observed among non-participants (see Table C.6 in Appendix C). If, as this finding suggests, programme participants were economically more disadvantaged than non-participants, comparing outcomes for participants and non-participants would probably understate the impact of the pilot on certain outcomes, like being in employment or school, as these outcomes are expected to be positively affected by individual/family wealth.

By identifying a comparison group (a subset of the non-participants group with similar characteristics to participants), the PSM method eliminates (or at least reduces) selection bias.

¹¹See Rosenbaum and Rubin (1983).

¹²A probit regression model was used to estimate the propensity scores. In this model, the dependent variable was a binary indicator denoting the individual’s participation status (the variable was coded 1 for participants and 0 for non-participants). The explanatory variables were observed determinants of participation, and were different depending on whether the analysis was conducted using survey or administrative data, reflecting the availability of different data.

¹³The number of matched non-participants used to produce the impact estimates in the analysis which employed administrative data was generally different from the number of participants. This is due to the fact that the matching process was implemented with replacement (this implies that a non-participant individual can be used as a match for more participants), and using the ‘ties’ option. The ties option uses all matched non-participants with the same propensity score (rather than only one of them) in the calculation of *impact estimates.

2.6 Application of the PSM to the impact analysis

The PSM method was used to select, among the 1,537 non-participants who responded to the survey, those most similar to the 908 survey respondents who participated in the pilot. Then, for each of these 908 participants and 1,537 non-participants, we estimated the probability of being recruited for the IF pilot (the propensity score) conditional on individual characteristics sourced from both the NPD and IF survey (these are listed in Tables C.1, C.2 and C.3 - see Appendix C). The matching approach resulted in 891 of the 908 participants being matched to 891 of the 1,537 non-participants, and these were the samples used to estimate survey-based impacts.

For the impact analysis which used administrative data, the probability of being recruited for the pilot was estimated, separately, for each of the three cohorts of starters (this means that a different set of propensity scores was obtained for each cohort). The individual determinants of selection into the IF pilot included the NPD/ILR variables used to profile participants and non-participants (see Table C.6 in Appendix C). The selection model also included six binary indicators denoting all possible combinations of months and years in which participants started (or pseudo-started, for non-participants) the IF pilot project, which varied depending on the cohort observed.¹⁴ Across the three cohorts of starters, 2,102 participants were matched to 2,023 non-participants (these are the numbers for educational outcomes, figures for behavioural outcomes were lower due to only two cohorts being observed).

2.7 SROI analysis

The SROI approach has several similarities to the economic method known as 'cost-benefit analysis' (CBA), and in practice the differences may be relatively limited. However, CBA is generally used for evaluation of larger government projects, whilst SROI tends to be used to look at social enterprises and their wider impact. The SROI approach aims to assess whether the benefits produced by projects, including the wider intangible benefits, exceed their costs. The SROI model outlines a methodology for calculating value as well as prescribing a set of principles for the evaluation of costs and benefits. There are different ways of monetising the benefits that occur, but including benefits to individuals and to society, as well as direct fiscal consequences of any programme. The costs are based on accounting data, which demonstrate the flows of funds from DWP to the contractors.

2.8 Data considerations

A key aim of PSM is to balance the participant and non-participant groups according to a suite of characteristics that are related to the outcomes being studied. Imbalances between the two groups may reflect differences in personal characteristics that influence young people to participate in the programme and/

¹⁴For example, for the cohort of starters in the academic year 2012/13, the six indicators denoted whether the young person started in November 2012, December 2012, January 2013, February 2013, March 2013 or April 2013. For non-participants, these variables indicated the month and year of the assigned pseudo start date.

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

or reflect selection criteria applied by the providers. Ideally, all important variables relating to differential selection effects between IF participants and non-participants would be adjusted for in the matching.

The variables available on the NPD for the matching exercise are plausibly related to the outcomes but there is a possibility that they do not capture certain influential selection criteria. Indeed, the matching diagnostics suggested that the matching procedure was not as successful as desired and that impact estimates may be biased as a result of this.¹⁵ The IF pilot was aimed at NEET young people and those at risk of becoming NEET and the NPD variables included in the PSM attempted to capture this aspect. Unfortunately, it is not possible empirically to assess the potential effect of failing to include variables capturing programme participation. It is plausible to assume that the individual-level NPD variables used to implement the PSM (notably those capturing persistent absenteeism and exclusion from school) partly controlled for these unobservable characteristics.

Another important consideration is that the two measures employed by this impact evaluation to capture attendance/behaviour-related outcomes (namely, whether pupils had been persistently absent or whether they had been excluded from school in a given academic year) did not match closely the definitions provided by the IF guidance¹⁶ Someone who had been excluded from school in the course of a given academic year could still have experienced, and demonstrated, a positive behavioural change for 13 continuous weeks, resulting in the behavioural outcome being achieved for the purpose of the IF payment. Similarly, it is possible that a pupil who had initially missed 15 per cent of their school's sessions (defined as persistently absent for the purpose of this evaluation) had later attended school for 13 consecutive weeks, proving they were no longer persistently truant. Therefore, the attendance/behaviour-related outcomes used here reflected more stringent definitions than those entailed by the IF guidance. This was due to the limited amount of exclusion-related measures available, and the decision that the persistent absenteeism data represented the best fit given the limited measures available. As a consequence, using these measures is likely to have resulted in understating the impact of the IF pilot, and estimates should be interpreted carefully. Programmes such as the IF often have a wide range of effects, over and beyond the specific outcomes that form the payment schedule. For example, there may have been gains in confidence and time management among those involved in the intervention, which represent tangible gains but do not trigger financial payments to providers. In this evaluation the data collection was largely restricted to the key outcomes explored here, and hence the full range of measures that might have been included in a full SROI analysis were not available. Had there been access to such data, the size of the benefits could well have been larger, although the impact study cautions against attributing a lot of changes to the programme itself.

¹⁵Appendix E presents the distributions of the propensity scores for IF participants and non-participants – see Figures E.1 to E.7)

¹⁶According to the IF guidance, changes in both persistent truancy and behaviour should be demonstrated for a minimum of 13 continuous weeks for the outcome to be achieved and the payment claimed.

The guidance refers to Section 91 of the Education and Inspection Act 2006, which defines a pupil having poor behaviour as someone 'whose behaviour is unacceptable, who break school rules or who fail to follow a reasonable instruction.' It also defines persistent truancy as 'deliberate absence for days or weeks at a time,' adding that 'This has to decrease to attendance levels associated with the average student.'

3 Results from the impact analysis based on survey data

3.1 Introduction

The Innovation Fund (IF) survey covered young people involved with all Round One projects delivered in England and a comparable, pre-selected group of non-participants, similar to participants with respect to a range of demographic and educational characteristics. IF participants who responded to the survey were aged between 14 and 18 at the start of the IF pilot projects, and this age range meant that the impact analysis which used survey data covered both young people who were NEET (participants aged 16 or more) and those at risk of becoming so (younger participants, aged 14 and 15). As the IF pilot was aimed at preventing young people from becoming NEET and attempted to re-engage those who were already NEET with employment, education and training, this evaluation sought to explore the impact of the programme on the likelihood of participants taking part in a range of activities related to employment, education and training.

As explained in Chapter 2, the Propensity Score Matching (PSM) method was used to refine the available group of non-participant survey respondents to further improve their similarity with the 891 participants who responded to the survey. The resulting comparison sample of 891 non-participants was used to produce the estimates of the impact of the IF pilot. Information about young people's current activities was collected through the survey, providing the individual outcomes needed to estimate the impact of the IF on the likelihood of participants being in activities related to employment, education and/or training around one year after they started on the pilot.¹⁷ This chapter reports the estimates of the impact of the IF pilot on the outcomes explored. It also offers possible ways of interpreting these estimates and briefly discusses data limitations that could have possibly affected the validity of the results.

3.2 Impact results

This section shows the estimated impact of the IF pilot on the likelihood of participants being involved with each of five activities (among school/college, paid job, training course/scheme, apprenticeship and traineeship) at the time of the interview. It also reports the estimated impact on participants' likelihood of being in employment, education or training (EET). The impact on the EET status is also shown by age of participants at IF start.

¹⁷Survey respondents who participated in the IF pilot were asked about their experiences of taking part in the projects and of receiving other support beyond the pilot. Questions about participation in programmes other than the IF pilot and receipt of other forms of support were also asked to non-participants. However, the low response to these questions (see tables in Appendix D) meant that helpful contextual information, potentially available to enhance the interpretation of the findings from the impact assessment, was considered unreliable and therefore not used.

3.2.1 Impact on current activities

The impact of the IF pilot was estimated as the difference between the proportions of the participant and comparison sample who reported taking part in each of five activities at the survey interview date, which happened approximately one year after participants started on the pilot. Survey respondents were asked whether they were currently in school or college, in a paid job, in a training course or scheme, in apprenticeship or in traineeship. They could choose more than one answer (i.e., mention one or more of these activities).

Table 3.1 shows the proportions of the participant and comparison sample who mentioned taking part in each of the five activities, and the impacts estimated based on these proportions. Only estimates which were statistically significant at the five per cent (or a more stringent) level were considered as supportive of the existence of an impact. Note that the figures in each of the first two columns of the table do not necessarily sum to 100 as respondents could report multiple activities (the activities reported are not mutually exclusive).

Table 3.1 Impact of the IF pilot on the likelihood of being in each activity among school/college, paid job, training, apprenticeship and traineeship at the time of interview

	Participants (%)	Comparison (%)	Impact (ppts)
School or college	68.1	77.9	-9.8***
Paid job	10.8	15.8	-5.0***
Training course/scheme	14.1	10.6	3.5**
Apprenticeship	7.1	4.6	2.5
Traineeship	6.6	4.6	2.0

Base: 1,782 respondents (891 participants and 891 comparators) for all outcome variables but Apprenticeship (607 participants and 606 comparators) and Traineeship (640 participants and 634 comparators).

**** and **: Statistically significant at the one and five per cent level, respectively.*

As Table 3.1 indicates, the proportion who reported to be in school or college at the time of the interview was lower among IF participants (68.1 per cent) compared to non-participants from the comparison group (77.9 per cent).¹⁸ The impact of the IF pilot on school participation, calculated as the difference between these two proportions, was -9.8 percentage points (ppts). A negative impact of around 10ppts means that 68 in every 100 participants were in school after starting the IF pilot but 78 in every 100 would have been in school had young people not participated in the programme (this is the counterfactual outcome estimated based on the comparison group). Basically, this finding suggests a negative impact of the IF pilot in that it indicates that, following participation in the programme, the proportion of young people attending school/college decreased.

¹⁸Individuals were advised that they could report being in school or college if they were on study leave or doing exams.

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

Survey respondents were also asked whether they were currently doing any kind of paid job, and were advised that by paid job it was meant any job where they regularly worked for four hours or more every week. Among IF participants, 10.8 per cent reported that they were doing such a job while a higher proportion (15.8 per cent) was observed among those in the comparison sample. These figures suggest that the IF pilot reduced the proportion of participants who were in employment by 5ppts. This impact figure means that 11 young people in every 100 who participated in the IF pilot were in a paid job around one year after starting the programme but there would have been 16 employed in every 100 in the absence of the pilot.

The proportion who reported to be in a training course or scheme was higher among IF participants (14.1 per cent) than their matched counterpart (10.6 per cent). Therefore, the estimated impact of the pilot on training participation was 3.5ppts: while approximately 14 in every 100 students were on training courses or schemes around one year after starting the IF, a lower number (11 in every 100) would have been on training courses at the same point in time had participants not been supported by the programme. Put simply, this finding suggests a positive impact of the IF pilot on training attendance as the proportion of participants who were in training increased as a result of the programme.

In the survey, young people were asked if they were currently doing an apprenticeship, and it was explained to them that an apprenticeship trains young people for a specific job while they are paid a regular wage or training allowance. The proportion of IF participants who mentioned that they were doing an apprenticeship at the time of the interview was 7.1 per cent, and this figure was higher than the corresponding figure reported among respondents from the matched comparison group (4.6 per cent). Although this finding suggests that the IF pilot increased the proportion of participants who were doing an apprenticeship by 2.5ppts, this estimate was not found to be statistically significant. A similar finding was obtained for the traineeship-related outcome.

3.2.2 Impact on the EET status

When interpreting the statistics in Table 3.1, it is important to note that survey respondents could have reported multiple current activities. For example, it is possible that by the time that young people were interviewed they had started a job and were at the same time undergoing a training programme or doing a traineeship. The impact estimates reported above are therefore helpful to explore whether (and to what extent) the aforementioned activities were enhanced by the IF pilot. However, they do not provide an indication of the number of young people affected by the programme.

As mentioned before, the IF pilot was introduced with the chief aim to reduce the number of young people who were either NEET or at risk of becoming NEET. Therefore, it was decided to use a more aggregated outcome measure, summarising the extent to which the IF was successful in achieving this goal. The impact of the pilot on the proportion of young people who were in Employment, Education or Training (EET) at the time of the interview was then explored, and the statistical results are illustrated in Table 3.2. A young person was considered as having the EET status if, when asked about their current activities, they reported that they were doing at least one among school/college, paid job, training, apprenticeship or traineeship.

Table 3.2 Impact of the IF pilot on the likelihood of being in Education, Employment or Training (EET) at the time of interview

	Participants (%)	Comparison (%)	Impact (ppts)
In EET	78.6	86.4	-7.9***

Base: 1,782 respondents (891 participants and 891 comparators).
 ***: Statistically significant at the one per cent level.

As we can see from the table, at the time of the survey interview 78.6 per cent of IF participants and 86.4 per cent of the matched non-participant sample qualified as being EET. These figures translate to an IF impact of -7.9ppts, suggesting that the proportion in EET reduced by eight in every 100 participants following participation in the pilot.

3.2.3 Impact by age of participants at IF start

The impact of the IF pilot was also explored by the age of IF participants at the start of the programme. The by-age impact estimates are reported in Table 3.3.

Table 3.3 Impact of the IF pilot on the likelihood of being in Education, Employment or Training (EET) at the time of interview, by age of participant at IF start

Age at IF start	Participants (%)	Comparison (%)	Impact (ppts)
14	97.6	99.6	-2.1**
15	90.5	91.5	-0.9
16	62.6	85.1	-22.5***
17	58.0	68.5	-10.4
18	54.3	64.7	-10.4

Base: 1,782 respondents (891 participants and 891 comparators). Of these, 288 participants and 280 comparators were aged 14, 211 participants and 201 comparators were aged 15, 168 participants and 166 comparators were aged 16, 130 participants and 131 comparators were aged 17, and 102 participants and 105 comparators were aged 18.

*** and **: Statistically significant at the one and five per cent level, respectively.

The findings indicated that, after approximately one year following their start on the IF pilot, younger recruits (14-year-olds) and those aged 16 at the time they started on the pilot were adversely affected by the programme. The IF was found to have reduced the proportion who reported being in EET by 2.1ppts among young people in the former group, and by 22.5ppts among those in the latter group.

Impacts were also estimated, separately, for younger participants (14- and 15-year-olds) and older participants (aged 16 or over). The impact estimates are reported in Table 3.4.

Table 3.4 Impact of the IF pilot on the likelihood of being in Education, Employment or Training (EET) at the time of interview, younger versus older starters

Age at IF start	Participants (%)	Comparison (%)	Impact (ppts)
14-15	94.7	96.1	-1.4
16-18	59.0	74.5	-15.5***

Base: 1,782 respondents (891 participants and 891 comparators). Of these, 489 participants and 491 comparators were aged 14 or 15, and 402 participants and 400 comparators were aged 16 or over.

***: Statistically significant at the one per cent level.

A statistical test was implemented to assess the existence of a difference between the impact observed for younger participants (-1.4ppts) and the impact detected for older participants (-15.5ppts). The positive difference between the two impact estimates (14ppts) was found statistically significant at the one per cent level, suggesting that the IF pilot was more detrimental for older than younger participants in its attempt to engage/re-engage individuals with employment, education and training. As indicated by the qualitative evaluation of the IF pilot (Thomas *et al.*, 2016), young people aged 16 and over tended to have more complex and multiple needs and barriers than younger participants, and were generally much harder to engage and to support. Therefore, the finding of smaller impacts for this age group compared to younger participants is not surprising.

3.3 Discussion of findings

The impact analysis which used survey data indicated that, at an early stage (i.e., around one year after starting on the IF pilot), the proportion of programme participants who were in school/college or in a paid job decreased as a result of the pilot. However, the IF was also found to have increased the proportion of young people in training. Overall, when the impact on these three activities was considered jointly, the pilot was found to have had a negative impact on the proportion of those who were in employment, education or training. This result was more accentuated for older IF participants (those aged 16 or over) than younger ones (14 and 15 year olds). The finding of a detrimental impact of the IF pilot on participants’ education- and employment-related outcomes is surprising, especially considering the positive findings of the qualitative evaluation report (Thomas *et al.*, 2016).

As stated previously in chapter 2, there are a number of data limitations which need to be taken into account when considering the above findings. It is possible that the impact estimates were influenced by the matching process which (due to a lack of data) was not able to control for all factors simultaneously affecting participation in the IF pilot and the outcomes explored. These factors may include personal characteristics and circumstances associated with a greater probability of being or becoming NEET, and motivation. It is also worth noting that the impact estimates presented in this chapter were simply based on the comparison of *post-programme* outcomes for the IF participant sample and a matched, refined comparison group of non-participants. Whether young people had been working, studying or in training before the start of the programme could not be observed, and this might have biased impact estimates. For example, if a much larger proportion of young people in the

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

matched comparison group were studying compared to the IF participant sample, it is possible that the pilot increased the proportion of those in school but not enough to close the initial gap.¹⁹

Another important aspect to bear in mind is that vulnerable and disadvantage young people could be involved with the programme for up to three years, in consideration of the challenging objective to address their complex needs. It is plausible that it took more than one year (the average time observed between the IF start and time of the interview) for the pilot to exert its beneficial effects on its participants. Therefore, in order to capture the benefits of the IF pilot, individual outcomes may need to be observed over a longer time span than it was possible in the survey-based impact evaluation.²⁰ This was possible within the context of the impact evaluation based on administrative data, which provided an opportunity to explore the existence of lagged IF effects (see next chapter).

Finally, it is possible that IF participants have received less non-IF related support compared to their matched counterpart. More support from other programmes similar to the IF (e.g., the Youth Engagement Fund and Youth Employment Initiative) could have helped non-participants achieve better outcomes compared to participants. Unfortunately, this hypothesis is difficult to test as the timing of alternative programmes and/or support to the IF is not exactly known, and the low response to the question about non-IF support (especially among IF participants – see Tables C.10 to C.18 in Appendix C) makes it difficult to make a final judgement about whether this is a plausible explanation of the negative impact estimates detected by the survey-based impact analysis.

¹⁹Many of the older IF participants (those aged 16 or over) were NEET at the start of the pilot as this was a pre-requisite for joining some projects (e.g., the Nottingham City Council project, which included one fourth of the individuals in the participant sample used to produce impact estimates). It is therefore possible that the failure to control for the selection criteria for inclusion in the IF pilot (including individuals' pre-treatment NEET status) resulted in the obtainment of a comparison group which was less of disadvantaged as (and therefore did not fully resemble) the participant sample.

²⁰Unfortunately, the second wave of the survey (which had the aim of exploring longer-term impacts than those reported here), had a very low response rate and therefore responses could not be considered reliable.

4 Results from administrative data:

Overall impact

4.1 Introduction

This chapter presents the estimates of the impact of the Innovation Fund (IF) pilot obtained following the analysis of administrative data. Similar to the impact analysis which used survey data, impact estimates were achieved using a comparison group of non-participant young people identified by means of the Propensity Score Matching (PSM) method. However, unlike the survey-based impact analysis, which included participants aged 14 or over at IF start, the impact analysis which used administrative data covered only younger participants, aged 14 or 15 at programme start (as explained previously, the focus on this age band was determined by the reality of the programme, which supported mainly those under the age of 16, and the need to compare the impacts of Round One and Round Two projects). Furthermore, the impact assessment which used administrative data observed impacts over a longer follow-up period (up to three-year after starting on the programme) than the survey-based impact analysis (which explored impacts one year after).

The impact evaluation based on administrative data explored the causal effect of the IF pilot on three educational outcomes, namely the proportions of participants who achieved each qualification among a first NQF level 1, 2 and 3 (due to data limitations, the achievement of a first Qualifications Qualifications and Credit Framework (QCF) Entry-level qualification could not be explored), and two attitudinal/behavioural outcomes, that is, the proportions of participants who were recorded as being persistently absent and excluded from schools, respectively. For each of the three cohorts of participants (programme starters in the academic years 2012/13, 2013/14 and 2014/15), impacts on educational outcomes were observed at the end of the academic year in which participants started the programme and at the end of each of the three subsequent years (the actual follow-up period depended on the cohort explored). The impact on behavioural outcomes was observed over the year following the academic year in which participants started the programme and (due to data limitations) could be assessed only for participants from the 2012/13 and 2013/14 cohorts.

The estimates reported in this chapter refer to the 'overall' impact of the IF pilot to distinguish them from the by-subgroup impact estimates, which are illustrated in the next chapter. The overall impact of the pilot is the effect of the programme on participant individuals, and as such it was calculated using information on all young people covered by the IF provision. A by-subgroup impact estimate refers instead to the effect of the programme on a specific subset of participants, and is obtained using only data on individuals comprised in this subset (e.g., in the by-gender subgroup analysis, the impact of the IF pilot on male participants is estimated using only data on participant boys).

4.2 Impact on educational qualifications

The impact of the IF pilot on the likelihood of achieving each of the three qualifications considered (first NQF level 1, 2 and 3) was estimated, separately, for participants from the 2012/13, 2013/14 and 2014/15 cohorts. For each cohort, a measure of the IF impact was obtained by contrasting the proportion of participants who experienced the outcome of interest (i.e., the proportion who achieved the specific qualification) with the proportion of young people in the comparison group who experienced the same outcome. The latter proportion (called the ‘counterfactual’) represents the proportion that would have been experienced among participants had they not been supported by the programme.

4.2.1 Impact on the likelihood of achieving a first NQF level 1 qualification

Table 4.1 shows, for each of the three cohorts, the proportions of young people in the participant and comparison samples who achieved their first NQF level 1 qualification at each time point between the end of the academic year in which they started (or pseudo started, for the comparison group) on the programme (time point 0) and three years after (time point 3).

Table 4.1 Proportions of the participant and comparison group who achieved a first NQF level 1 qualification at each time point between the end of the IF start year and three years after

	2012/13 cohort		2013/14 cohort		2014/15 cohort	
	Participants %	Comparison %	Participants %	Comparison %	Participants %	Comparison %
End of academic year (0)	23.2	24.8	15.4	14.0	28.9	26.1
One year after (1)	82.6	78.6	81.1	76.9	72.2	62.5
Two years after (2)	94.3	93.1	91.8	90.6		
Three years after (3)	95.2	93.1				

Base: 949 participants and 876 comparators (2012/13 cohort), 668 participants and 656 comparators (2013/14 cohort), and 485 participants and 491 comparators (2014/15 cohort).

Figures labelled as ‘End of academic year’ indicate proportions of young people who obtained their first NQF level 1 qualification between their IF start/pseudo-start date and the end of the academic year in which this date fell. For example, 23.2 per cent of IF participants from the 2012/13 cohort had achieved a first NQF level 1 qualification between their IF start date and the end of the academic year 2012/13 while a slightly higher proportion among those in the comparison sample (24.8 per cent) had experienced this outcome between their pseudo start date and the end of the academic year 2012/13.

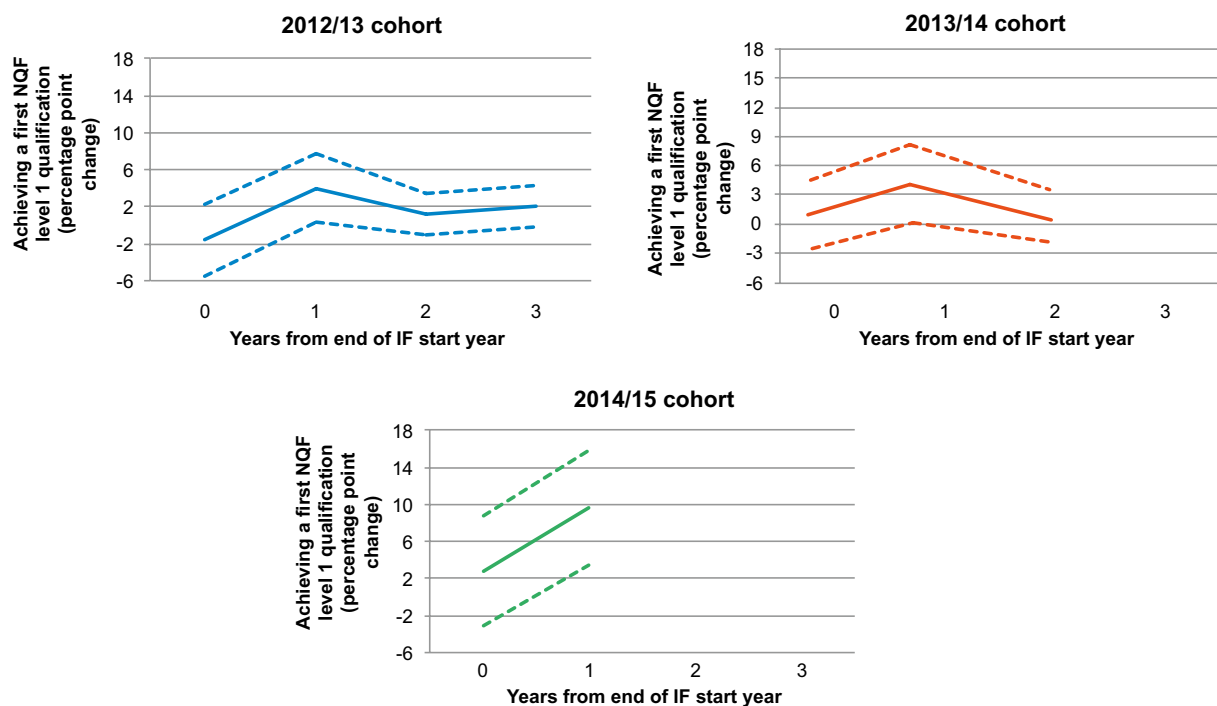
Similarly, figures labelled as ‘One year after,’ ‘Two years after’ and ‘Three years after’ denote proportions of individuals who obtained their first NQF level 1 qualification at some point in time between their IF start (or pseudo-start) date and the end of

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

the first, second and third academic year, respectively, following the academic year in which they started (or pseudo-started) the IF pilot. At each point in time and for each cohort, the difference between the proportions observed for the participant and comparison samples provides an estimate of the impact of the IF pilot on the likelihood of achieving a first NQF level 1 qualification.

Figure 4.1 provides a visual representation of the impact of the IF pilot on the likelihood of participants achieving a first NQF level 1 qualification. Separate impacts were estimated for each of the three cohorts of starters. For each cohort, impact estimates are shown together with their confidence intervals (upper and lower bounds), which are represented by means of dashed lines surrounding the impact trajectory. When both the lower and upper bounds of the confidence interval are positive (or negative) this means that the estimated impact was statistically significant at the five per cent level. Due to the large number of impact estimates produced, only those which were found statistically significant at this level are discussed in this chapter. (When considerations are made based on impact estimates non-statistically significant this will be explicitly stated.)

Figure 4.1 Impact of the IF pilot on the likelihood of achieving a first NQF level 1 qualification at each time point between the end of the IF start year and three years after



The most important findings that can be gauged by looking at the graphs can be summarised as follows:

- Around one year after participants from the 2012/13 cohort started on the IF pilot, the programme was found to have increased the proportion who achieved a first NQF level 1 qualification by four percentage points (ppts). This means that the proportion of participants who obtained this qualification (82.6 per cent, as indicated in the first column of Table 4.1 under the outcome ‘One year after’) was

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

4ppts higher compared to what it would have been had participants not started the programme (78.6 per cent, as the figure for the comparison group in the second column of Table 4.1 indicates).

- The proportion of participants from the 2014/15 cohort who achieved their first NQF level 1 qualification around one year after starting on the IF was found to have increased by nearly 10ppts as a result of the pilot: the proportion of achievers among participants was 72.2 per cent (as indicated in the fifth column of Table 4.1 under the outcome ‘One year after’) but would have been 62.5 per cent in the absence of the pilot (see the sixth column of Table 4.1).

The finding of a positive impact of the IF pilot on the likelihood of participants obtaining a first NQF level 1 qualification confirms the findings of the IF qualitative evaluation (Thomas *et al.* 2016). This reported that ‘Qualifications delivered by IF project staff were at entry-level or Level 1 and were seen as practical enhancements of young people’s CVs that would help to raise self-esteem and also be recognised as relevant by potential employers’ (Thomas *et al.*, 2016, page 45). The authors highlighted the importance of ‘moving young people forward who were not yet ready for employment, such as entry-level and Level 1 qualifications, work experience and voluntary work placements’ (Thomas *et al.*, 2016, page 36).

4.2.2 Impact on the likelihood of achieving a first NQF level 2 qualification

Table 4.2 shows, separately for each of the three cohorts, the proportions of the participant and comparison samples who achieved a first NQF level 2 qualification at all time points considered.

Table 4.2 Proportions of the participant and comparison group who achieved a first NQF level 2 qualification at each time point between the end of the IF start year and three years after

	2012/13 cohort		2013/14 cohort		2014/15 cohort	
	Participants %	Comparison %	Participants %	Comparison %	Participants %	Comparison %
End of academic year (0)	2.6	5.5	2.5	5.3	1.0	2.5
One year after (1)	15.0	29.3	15.4	26.4	12.4	20.2
Two years after (2)	32.2	46.6	21.1	39.0		
Three years after (3)	32.8	47.0				

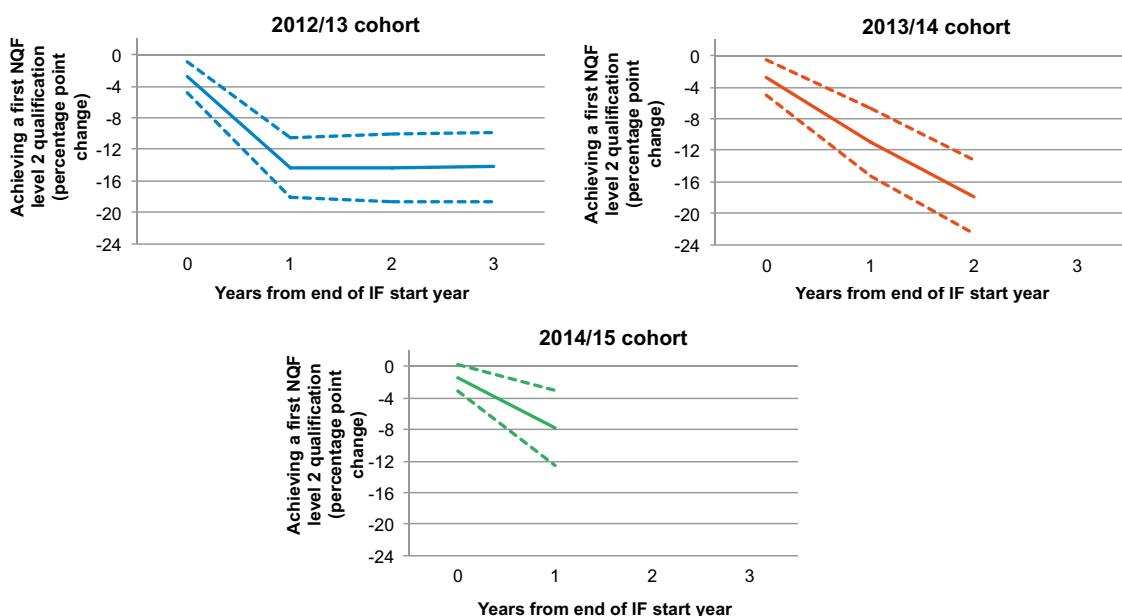
Base: 949 participants and 876 comparators (2012/13 cohort), 668 participants and 656 comparators (2013/14 cohort), and 485 participants and 491 comparators (2014/15 cohort).

The impact of the IF pilot on the likelihood of participants obtaining a first NQF level 2 qualification is represented graphically by means of Figure 4.2. As the figure shows, the following negative impacts were detected:

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

- By the end of the academic year in which they started on the IF pilot, participants from both the 2012/13 and 2013/14 cohorts had experienced a reduction of almost 3ppts in the proportion who achieved a first NQF level 2 qualification.
- At later time points, the impacts detected for participants from the 2012/13 cohort were still negative but much larger (in absolute value), being them in the region of -14ppts. Participants from the 2013/14 cohort experienced reductions of 11 and 17.9ppts one and two years after starting on the pilot, respectively.
- A negative impact was also found for participants from the 2014/15 cohort one year after starting on the IF: the programme was found to have reduced the proportion who obtained a first NQF level 2 qualification by 7.8ppts.

Figure 4.2 Impact of the IF pilot on the likelihood of achieving a first NQF level 2 qualification at each time point between the end of the IF start year and three years after



Similar to the negative impacts that the survey-based analysis detected on the likelihood of IF participants being in school/college and of being in a paid job, the finding of a negative impact on the likelihood of participants achieving their first NQF level 2 qualification could be attributable to the inability of the PSM methodology (due to data limitations) to control for unobserved individual characteristics highly predictive of both programme participation and the outcomes explored. It is also possible that a larger proportion of participants than observed would have obtained this qualification in the absence of the programme. In other words, the IF pilot might have prevented some participants from getting a qualification that they would have naturally attained.

As reported by Thomas *et al.* (2016), due to cash-flow requirements, many projects (especially the largest ones) turned away from the initially envisaged staged-progression delivery model to pursue a 'rolling intake' model, based on large numbers of recruits with fewer outcomes per person achievable in a short time frame (notably, NQF Level 1 qualifications and outcomes related to improvements in

school behaviour and attendance).²¹ Therefore, supporting participants towards the achievement of qualifications higher than NQF level 1 seems not to have happened (or have been limited) in reality. It is possible that, in order to achieve the cash flow needed to operate, some projects ‘traded’ individual targets defined in terms of higher-level qualifications with other non-educational targets (there is strong qualitative evidence of improved attendance and attitude towards school among programme participants).

4.2.3 Impact on the likelihood of achieving a first NQF level 3 qualification

The proportions of the participant and comparison samples who achieved their first NQF level 3 qualification, for each of the three cohorts and at all time points considered, are reported in Table 4.3. Note that proportions equal to zero indicate that nobody among participants or comparators achieved a first NQF level 3 qualification. Exception to this are comparison individuals from the 2013/14 cohort at the end of the academic year in which they pseudo-started the pilot and comparators from the 2014/15 cohort one year after, as in these instances zero values reflect negligible proportions who achieved the qualification in question (i.e., the proportions were approximated to zero because they were extremely small).

Table 4.3 Proportions of the participant and comparison group who achieved a first NQF level 3 qualification at each time point between the end of the IF start year and three years after

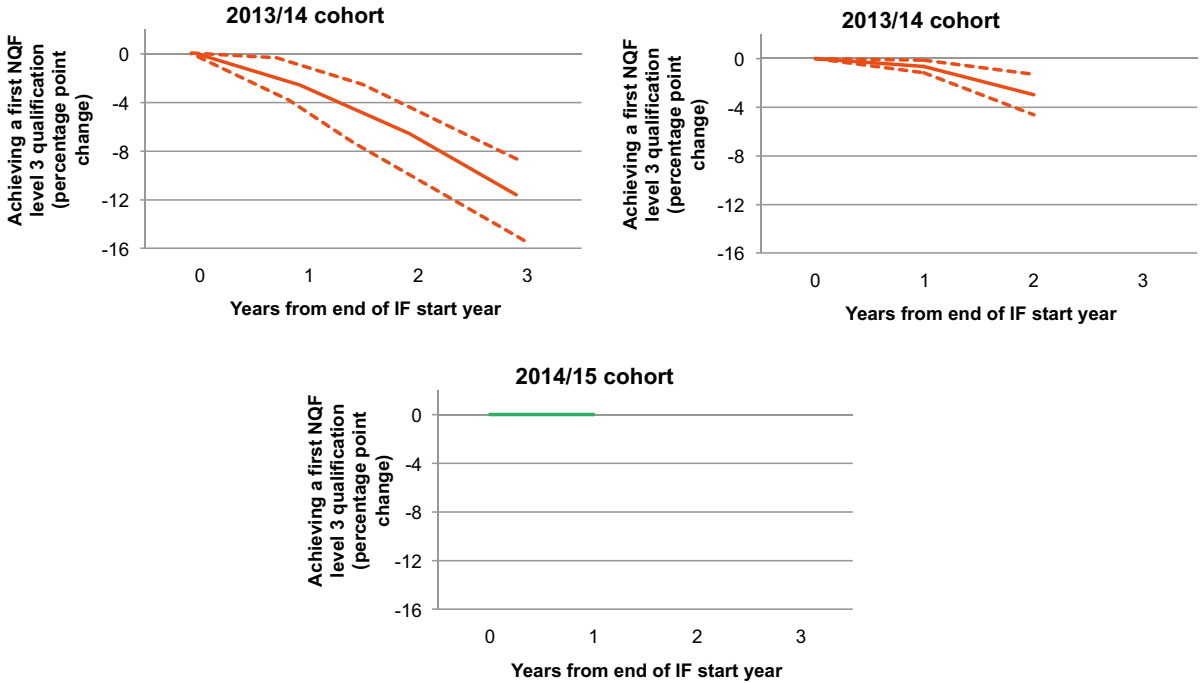
	2012/13 cohort		2013/14 cohort		2014/15 cohort	
	Participants %	Comparison %	Participants %	Comparison %	Participants %	Comparison %
End of academic year (0)	0.0	0.0	0.0	0.0	0.0	0.0
One year after (1)	0.8	2.7	0.0	0.6	0.0	0.0
Two years after (2)	5.8	11.5	0.6	3.6		
Three years after (3)	7.6	19.2				

Base: 949 participants and 876 comparators (2012/13 cohort), 668 participants and 656 comparators (2013/14 cohort), and 485 participants and 491 comparators (2014/15 cohort).

The impact of the IF pilot on the likelihood of participants obtaining a first NQF level 3 qualification is depicted in Figure 4.3. Note that the impact detected for the 2014/15 cohort one year after starting on the programme (not statistically significant) was positive but extremely small (close to zero), and this is the reason that the impact trajectory looks flat in the graph and the confidence interval is not clearly visible.

²¹As reported by Thomas *et al.* (2016), preparing young people for NQF levels 2 and 3 qualifications would have required a high number of learning hours, and this is the reason that no ‘in-house’ Level 2 qualifications were delivered by the projects.

Figure 4.3 Impact of the IF pilot on the likelihood of achieving a first NQF level 3 qualification at each time point between the end of the IF start year and three years after



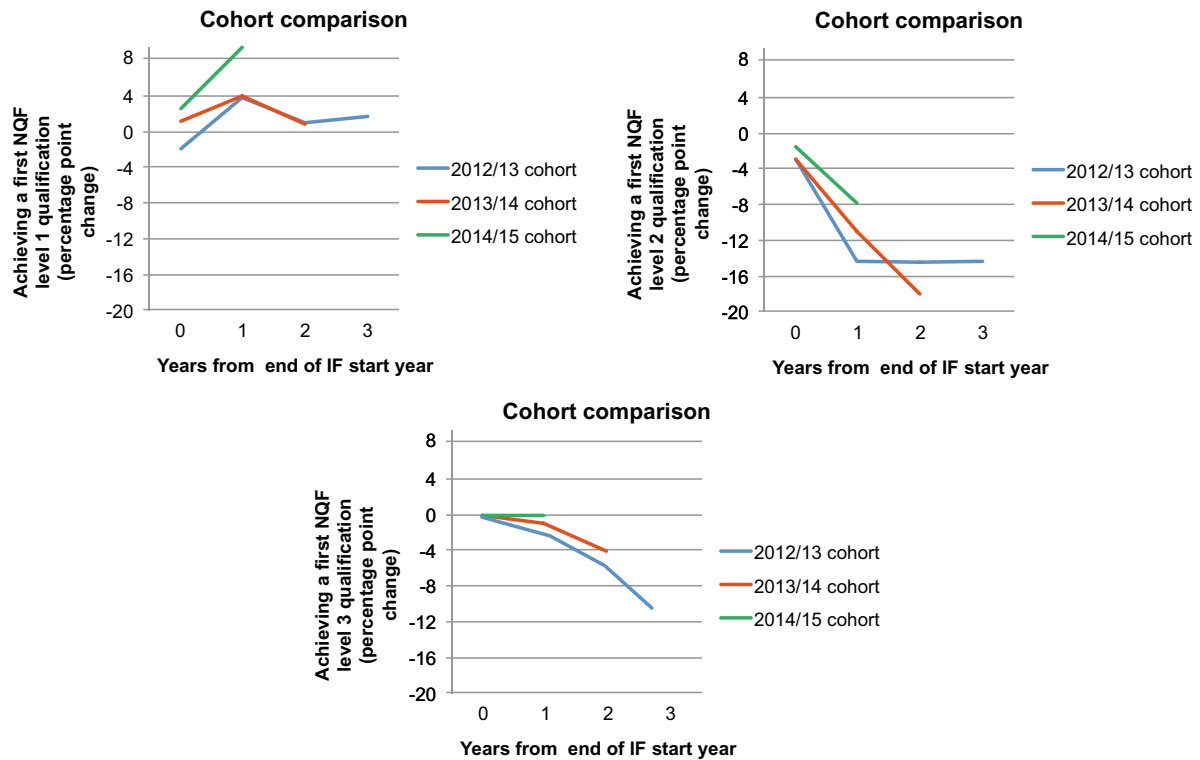
As Figure 4.3 clearly shows, evidence of a negative impact of the IF pilot was found. The main findings can be summarised as follows:

- Over time, the IF pilot progressively reduced the proportion of participants from the 2012/13 cohort who obtained a first NQF level 3 qualification. The reduction ranged from 1.9ppts one year after participants started on the programme to 11.6ppts three years after.
- Negative impacts, albeit smaller, were also detected for participants from the 2013/14 cohort one and two years after starting on the IF pilot (impact estimates were -0.6 and -3ppts, respectively).

4.2.4 Comparing impacts across cohorts

Figure 4.4 compares the impact trajectory for each educational outcome across the three cohorts explored. Essentially, the figure reports the impact estimates already shown in Figures 4.1 to 4.3 but without their confidence intervals, to facilitate the visual comparison of trends. The patterns observed seems to suggest that the positive impact of the IF pilot on the likelihood of achieving a first NQF level 1 qualification becomes larger, while at the same the negative impact on the likelihood of obtaining a first NQF level 2 or 3 qualification attenuates for later cohorts.

Figure 4.4 Impact of the IF pilot on the likelihood of achieving each qualification among a first NQF level 1, 2 and 3 at each time point between the end of the IF start year and three years after



The better performance (i.e., higher impact) generally observed for later cohorts compared to earlier ones could possibly reflect learning effects. That is, programme providers understanding how to manage difficulties encountered in the initial implementation stage, and later cohorts benefiting from past mistakes made by earlier cohorts. However, this consideration is based on a simple observation of impact trends, regardless of the statistical significance of the estimates and bearing in mind that sample sizes are smaller for later cohorts. Therefore, while findings are suggestive of possible learning effects, this is a merely speculative observation.

4.3 Impact on school attendance and behaviour

The estimates of the impact of the IF pilot on the likelihood of participants from the 2012/13 and 2013/14 cohorts being persistently absent and excluded from school, together with the proportions of the participant and comparison samples who experienced each of these two outcomes, are reported in Table 4.4.

Table 4.4 Impact of the IF pilot on the likelihood of participants being persistently absent and excluded from school in the academic year after starting the pilot

	2012/13 cohort	2013/14 cohort
Persistently absent (impact in ppts)	5.3 ***	4.0**
Participants (%)	(11.9)	(10.0)
Comparison (%)	(6.6)	(6.0)
Excluded (impact in ppts)	5.0**	4.2**
Participants (%)	(16.4)	(13.5)
Comparison (%)	(11.5)	(9.3)

Base: 664 participants and 635 comparators (Persistently absent), and 706 participants and 659 comparators (Excluded) for the 2012/13 cohort; 510 participants and 490 comparators (Persistently absent), and 547 participants and 532 comparators (Excluded) for the 2013/14 cohort.

*** and **: Statistically significant at the 1 and 5 per cent level, respectively.

The findings suggested that the IF pilot adversely affected participants' school attitudes and behaviours in that:

- In the academic year after they started on the IF pilot, participants from the 2012/13 cohort experienced a (statistically significant) increase of 5.3ppts in the proportion recorded as persistently absent. The IF increased the proportion of individuals in this cohort who received one or more exclusions by a similar figure.
- The proportion of participants from the 2013/14 cohort who were persistently absent from school increased by 4ppts (statistically significant) as a consequence of the IF pilot, and a similar impact was detected on the proportion of the same cohort who were excluded from school.

The finding of an adverse effect of the IF pilot on participants' school attendance and behaviour is surprising as it conflicts with the reports of the qualitative impact evaluation (Thomas *et al.*, 2016). This highlighted the schools' positive reports of young people's improved attitudes, attendance and behaviour resulting from the support that they received through the IF projects. A possible explanation for the divergent findings is to be found in the outcome measures used in this quantitative evaluation. As previously explained in section 2.8, these did not closely match the definitions of attitudinal and behavioural outcomes paid for by the IF pilot and, as a result of this, they are likely to have resulted in underestimating the impact of the IF pilot on both outcomes.

Furthermore, the results of the covariate balancing tests (not provided in this report) indicated that, for some of the impacts reported in Table 4.4, the comparison group used to produce the estimates was not perfectly balanced in terms of all individual characteristics used to define similarity with the participant sample. For example, the comparison groups used to estimate the 4ppts impact on the likelihood of being persistently absent from school for the 2013/14 cohort included a lower proportion of individuals with special education needs compared to participants (reflecting in a 12.4 per cent bias figure), and the 5ppts impact on the likelihood of being excluded for the 2012/13 cohort had a higher proportion of individuals recorded as persistently absent (resulting in a -14.4 per cent bias).

The above considerations suggest that caution should be exercised when using the impact estimates presented in this report to judge the extent to which the attitudinal and behavioural outcomes paid for by the pilot were achieved.

4.4 Further considerations

An important consideration concerns the comparison groups used to produce the impact estimates. While, overall, the covariate balancing tests indicated that the matching process was successful in identifying groups of non-participants similar to participants (in terms of all the baseline characteristics considered), the very low values of the estimated propensity scores indicated that none of the predictors of programme participation used to specify the score were very powerful. This suggests that some important determinants of participation have not been included in the analysis. These include individual characteristics and circumstances associated with a greater risk of becoming NEET (e.g., being a teenage parent, drug or alcohol misuser, young carer or gang member), and personal motivation. The impossibility to control for these important aspects (due to data limitations) suggests that the impact estimates shown in this report should be considered with caution.

Another aspect to consider is whether the findings reported in this chapter can be extended to the entire population of IF participants. As discussed in section 2.2.2, the cohort approach used meant that the sample used to produce impact estimates was limited to participants from the 2012/13, 2013/14 and 2014/15 cohorts (all November-April starters). Sensitivity checks were conducted to explore whether the inclusion in the analysis of participants who started the IF pilot in the months of September and October of the three academic years considered altered the impact estimates. The results indicated that effect of expanding the participant sample to include these individuals varied depending on the cohort explored. For example, larger positive impacts on the likelihood of achieving a first NQF level 1 qualification were obtained for the 2012/13 cohort two and three years after the IF start year, while a smaller (positive) impact was detected for the 2014/15 cohort one year after. However, the general results (positive impact of the pilot on the likelihood of participants achieving their first NQF level 1 qualification, and negative impact on the likelihood of achieving higher-level qualification) remained unaffected.

5 Results from administrative data:

Subgroup impacts

5.1 Introduction

Within the context of the impact evaluation based on administrative data, two by-subgroup impact analyses were conducted. These explored the impact of the Innovation Fund (IF) pilot by

- Round of commissioning: The by-round impact analysis investigated, separately, the impact of the programme for participants in Round One and Round Two projects.
- Gender of the participants: The by-gender impact analysis explored the effect of the IF pilot, separately, for male and female participants.

Beyond assessing the effect of the IF pilot for each subgroup of interest, the by-subgroup impact analyses were concerned with testing whether the differences in impacts between subgroups (i.e., between participants in Round One and Round Two projects, and between male and female participants) were statistically significant. The finding of statistically significant differences would suggest that the pilot were more effective for one subset of individuals than for another in achieving the outcomes set by the programme.

5.2 By-round impact

The by-round subgroup impact analysis estimated the impact of the IF pilot, separately, for young people who participated in any of the Round One projects (Triodos, APM Ltd, Links to Success and PEF) and for those participating in Round Two projects (Prevista, Teens & Toddlers and Energise Ltd).

While in the IF pilot's initial implementation stage, the provision of the Round One projects covered young people aged between 14 and 24 (therefore including both individuals who were NEET and those at risk of becoming NEET), there was a later shift to an early intervention (preventative) model which resulted in supporting mostly school-age children (see Thomas *et al.*, 2016). This is the main reason that this evaluation focused exclusively on participants aged 14 and 15. Furthermore, the Round Two projects were focused purely on those aged 14 and 15 (at risk of becoming NEET), and therefore estimating impacts for participants within this age range also for Round One projects enhanced the comparability of the findings from the two rounds.

This section reports the estimates of the impact of the IF pilot on educational and school attendance/behaviour-related outcomes for participants in Round One and Round Two projects.²² Aside from assessing the impact of the programme for each round of commissioning, the focus of the by-round subgroup impact analysis was

²²The proportions of the participant and comparison groups used to obtain the impact estimates presented in this chapter are reported in Appendix F (see Tables F1 to F4).

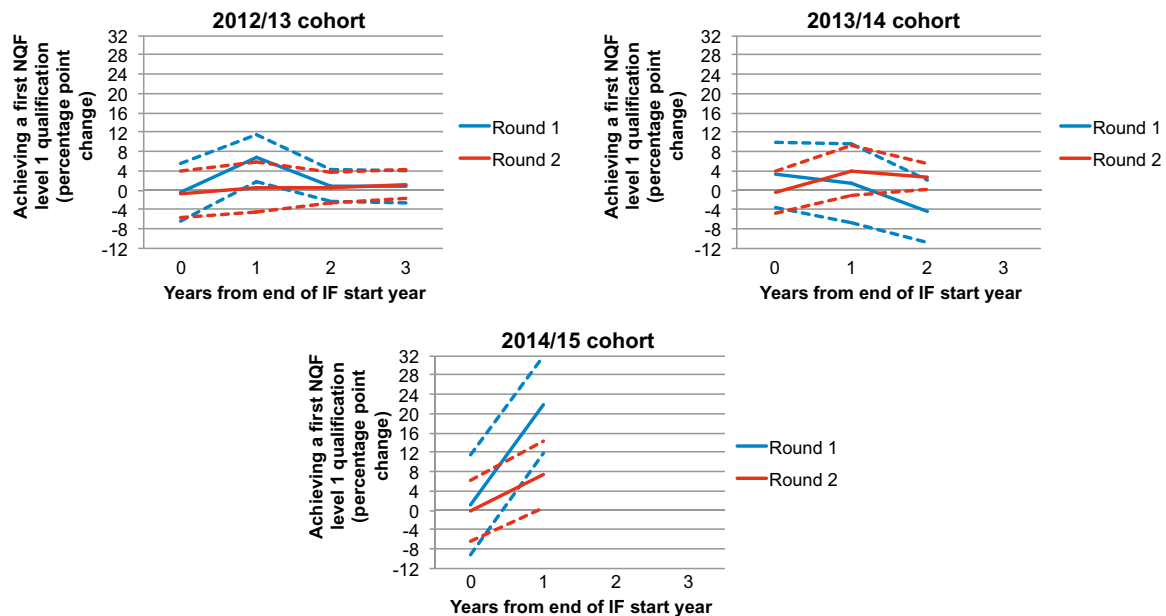
Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

testing whether the impact experienced by participants in the Round One projects was higher (or lower) than the impact on participants in the Round Two projects. The results of these tests are illustrated in this section.

5.2.1 Impact on the likelihood of achieving a first NQF level 1 qualification

The impact of the IF pilot on the likelihood of achieving a first NQF level 1 qualification was estimated, separately, for participants in Round One and Round Two projects. Figure 5.1 offers a visual representation of the by-round impact estimates for each of the three cohorts.

Figure 5.1 Impact of the IF pilot on the likelihood of achieving a first NQF level 1 qualification at each time point between the end of the IF start year and three years after, by round of commissioning



The main findings can be summarised as follows:

- Among young people from the 2012/13 cohort who started on the Round One projects, the proportion who obtained a first NQF level 1 qualification had increased by 6.7ppts around one year after their project start.
- Individuals from the 2014/15 cohort who started any of the Round One projects had experienced a 21.7ppts increase in the proportion who achieved a first NQF level 1 qualification one year after starting on the project. Participants in the Round Two projects experienced a positive increase of 7.4ppts at the same point in time.
- Two years after starting on the IF pilot, young people from the 2013/14 cohort who were supported by the Round Two projects had experienced a 2.8ppts increase in the proportion who obtained their first NQF level 1 qualification.

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

The results of the tests to assess the existence of differences in magnitude between the impact of Round One and Round Two projects on the likelihood of achieving a first NQF level 1 qualification are shown in Table 5.1. Note that the difference between the two subgroup impact estimates can be found statistically significant regardless of whether either of the two individual estimates is statistically significant.

Table 5.1 Difference between Round One and Round Two impacts on the likelihood of achieving a first NQF level 1 qualification

	End of year	One year after	Two years after	Three years after
2012/13 cohort	0.3	6.1**	0.4	-0.6
2013/14 cohort	3.7	-2.7	-7.1**	
2014/15 cohort	1.3	14.3**		

** : Statistically significant at the 5 per cent level.

As shown in the table the following findings emerged:

- One year after starting on the IF pilot, individuals from the 2012/13 and 2014/15 cohorts who were supported by the Round One projects had experienced a larger impact compared to young people from the same cohorts who participated in the Round Two projects. The proportion who achieved a first NQF level 1 qualification raised by 6.1ppts (2012/13 cohort) and 14.3ppts (2014/15 cohort) more for participants in Round One projects compared to participants in Round Two projects.
- A negative impact difference (-7.1ppts) for the 2013/14 cohort at the time point ‘Two years after’ suggested that Round One projects were less effective than Round Two projects in supporting young people towards attaining their first NQF level 1 qualification.

5.2.2 Impact on the likelihood of achieving a first NQF level 2 qualification

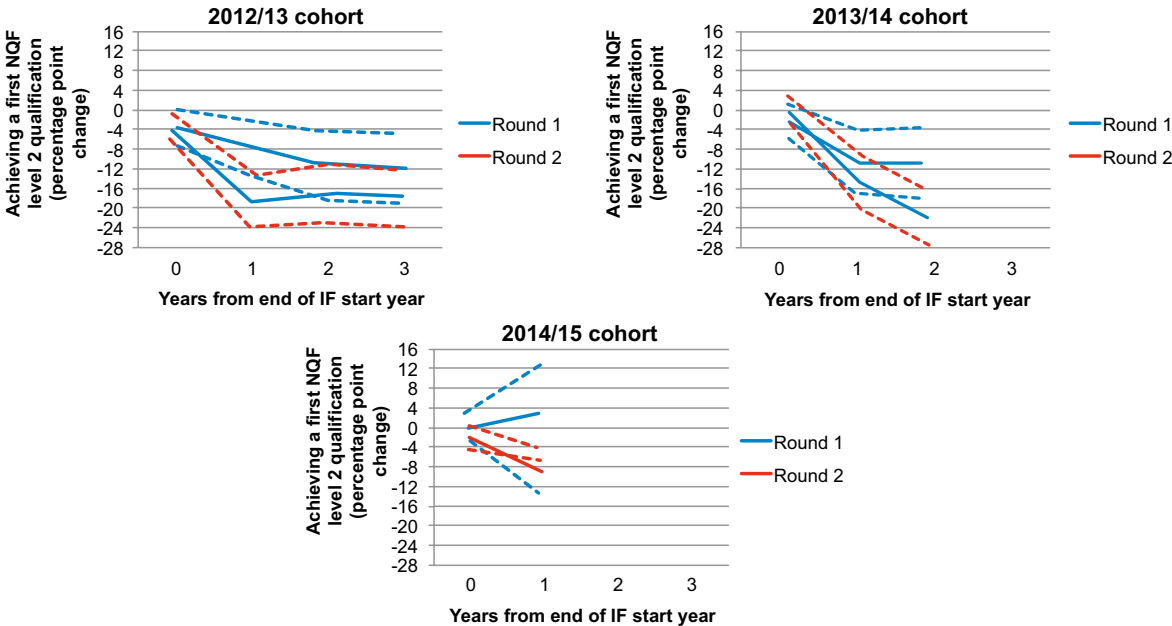
The findings from the by-round analysis of the impact of the IF on the likelihood of participants from the three cohorts achieving a first NQF level 2 qualification are illustrated in Figure 5.2. They suggested that:

- Starting either a Round One or Round Two project reduced the likelihood of individuals from the 2012/13 cohort achieving a first NQF level 2 qualification by some 3.5ppts, by the end of their IF start year.
- For young people from the same cohort, the negative impact increased (in absolute value) at later time points for both rounds. The impact for the Round One projects reached its lowest peak (-12ppts) three years after starting on the IF, while the lowest impact for the Round Two projects (-18ppts) was observed one year after.
- Negative impacts were also observed for the 2013/14 cohort at the time points ‘One year after’ and ‘Two years after.’ For example, two years after starting on the pilot, the proportion of Round One starters who achieved a first NQF level 2 qualification was reduced by 12ppts, while the impact observed for Round starters was -22ppts.

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

- Participants in the Round Two projects from the 2014/15 cohort experienced a 1.5ppts reduction in the proportion who achieved a first NQF level 2 qualification one year after starting on the projects, and a 9ppts reduction two years after.

Figure 5.2 Impact of the IF pilot on the likelihood of achieving a first NQF level 2 qualification at each time point between the end of the IF start year and three years after, by round of commissioning



The trends displayed in Figure 5.2 suggested that, in the main, Round One projects were more effective (although, in most cases, this actually meant less adverse) than Round Two projects in supporting young people towards the achievement of their first NQF level 2 qualification. Testing the existence of differences between Round One and Round Two impacts (see Table 5.2) confirmed this:

- For both the 2012/13 and 2014/15 cohorts, one year after starting on the programme, participants in Round One projects had experienced a smaller negative impact (by 10.4ppts) compared to participants in Round Two projects.
- A similar, positive impact difference (10.9ppts) was found for the 2013/14 cohort at the time point ‘Two years after,’ suggesting that by this point in time Round One projects had been less detrimental than Round Two projects in supporting young people towards achieving a first NQF level 2 qualification.

Table 5.2 Difference between Round One and Round Two impacts on the likelihood of achieving a first NQF level 2 qualification

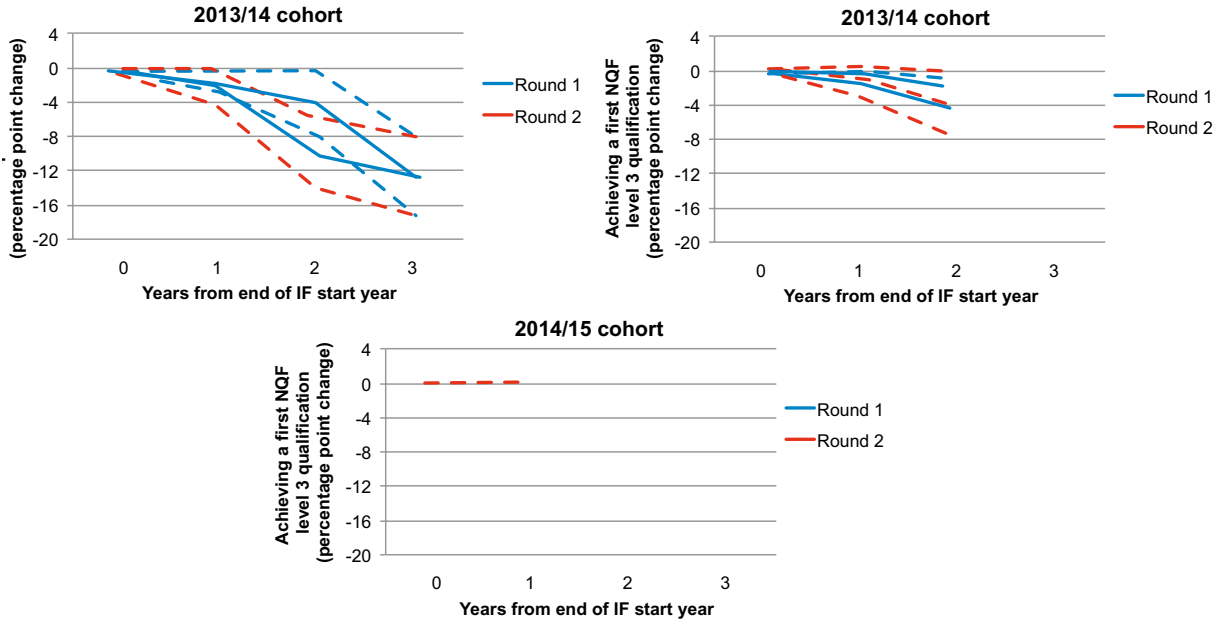
	End of year	One year after	Two years after	Three years after
2012/13 cohort	0.1	10.4***	5.8	6.2
2013/14 cohort	-1.8	3.3	10.9**	
2014/15 cohort	1.0	11.5**		

**: Statistically significant at the five per cent level.

5.2.3 Impact on the likelihood of achieving a first NQF level 3 qualification

The impact of each of the two rounds of commissioning on the likelihood of participants achieving a first NQF level 3 qualification is illustrated in Figure 5.3.

Figure 5.3 Impact of the IF pilot on the likelihood of achieving a first NQF level 3 qualification at each time point between the end of the IF start year and three years after, by round of commissioning



The statistical findings suggested that:

- Both Round One and Round Two projects reduced the likelihood of achieving a first NQF level 3 qualification for individuals from the 2012/13 cohort. For example, three years after starting on either round, the proportion who achieved this qualification was reduced by 12ppts.
- Negative impacts, although smaller (in absolute value), are observed for the 2013/14 cohort. For example, a 4ppts reduction in the proportion who achieved a first NQF level 3 qualification is observed among Round One starters.

As shown in Table 5.3, in relation to achieving a first NQF level 3 qualification the only statistically significant difference in impacts between the two rounds was the one observed two years after the IF start year for IF participants from the 2012/13 cohort. A negative impact difference of 5.8ppts suggests that the Round One projects were more detrimental than Round Two projects in supporting participants towards the achievement of their first NQF level 3 qualification.

Table 5.3 Difference between Round One and Round Two impacts on the likelihood of achieving a first NQF level 3 qualification

	End of year	One year after	Three years after	
2012/13 cohort	0.0	-0.5	-5.8 **	-0.2
2013/14 cohort	0.0	-0.9	-2.1	
2014/15 cohort	0.0	0.0		

** : Statistically significant at the five per cent level.

5.2.4 Impact on school attendance and behaviour

The by-round analysis of the IF impact on the likelihood of participants being persistently absent and excluded from school, respectively, detected some adverse effects. As shown in Table 5.4, the results indicated that:

- The IF pilot increased the proportion of Round One project starters from the 2013/14 cohort who were recorded as being persistently absent by 8ppts in the academic year after young people started on the projects.
- Participants in the Round Two projects from the 2012/13 cohort also experienced an increase (estimated to be 4.6ppts) in the proportion recorded as being persistently absent.
- Round Two projects increased the proportion of individuals excluded from school by 8.2ppts for starters from the 2012/13 cohort, and by 6ppts for starters from the 2013/14 cohort.

Table 5.4 Impact of the IF pilot on the likelihood of participants being persistently absent and excluded from school in the academic year after starting the pilot, by round of commissioning

	2012/13 cohort	2013/14 cohort
Round One		
Persistently absent (impact in ppts)	1.2	8.0**
Excluded (impact in ppts)	4.1	2.8
Round Two		
Persistently absent (impact in ppts)	4.6**	2.8
Excluded (impact in ppts)	8.2***	6.00***

Base (Round One): 250 participants and 221 comparators (Persistently absent), and 281 participants and 260 comparators (Excluded) for the 2012/13 cohort; 150 participants and 150 comparators (Persistently absent), and 178 participants and 176 comparators (Excluded) for the 2013/14 cohort.

Base (Round Two): 414 participants and 395 comparators (Persistently absent), and 425 participants and 410 comparators (Excluded) for the 2012/13 cohort; 362 participants and 356 comparators (Persistently absent), and 371 participants and 375 comparators (Excluded) for the 2013/14 cohort.

*** and **: Statistically significant at the one and five per cent level, respectively.

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

No statistically significant difference was found between the impacts estimated for the two rounds on the likelihood of IF participants being persistent absent and excluded from school, respectively.²³ This suggests that neither of the two rounds was more effective than the other in producing behavioural changes among programme participants.

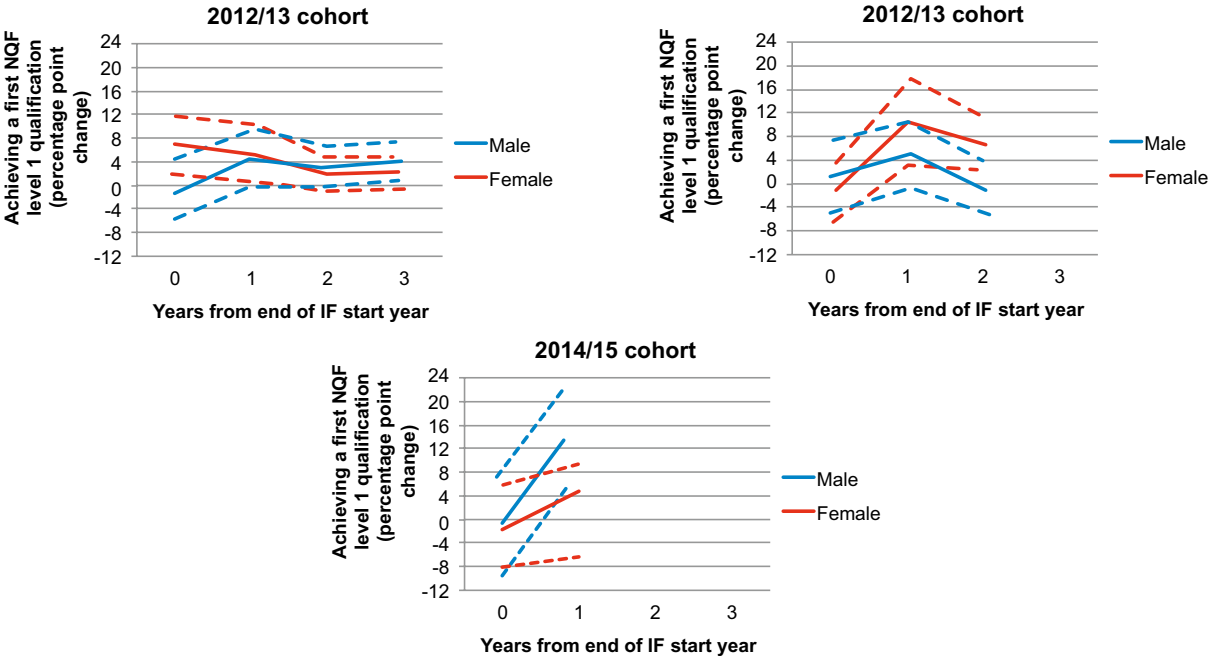
5.3 By-gender impact

The by-gender impact analysis was aimed at estimating the impact of the IF pilot, separately, for male and female participants. Additionally, it sought to explore whether the programme was more effective for boys than girls (or the other way around) in achieving the educational and behavioural outcomes of interest. There is no particular reason to believe that the programme was more helpful for participants of a specific gender compared to the other, and therefore the by-gender impact analysis was purely exploratory.

5.3.1 Impact on the likelihood of achieving a first NQF level 1 qualification

The impact of the IF pilot on the likelihood of achieving a first NQF level 1 qualification was estimated, separately, for male and female participants. The by-gender impact estimates are illustrated in Figure 5.4²⁴

Figure 5.4 Impact of the IF pilot on the likelihood of achieving a first NQF level 1 qualification at each time point between the end of the IF start year and three years after, by gender of participant



²³See Table F.5 in Appendix F.

²⁴The proportions of the participant and comparison group used to estimate impacts are reported in Table G.1 in Appendix G

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

The statistical findings suggested that:

- The IF pilot had a positive impact on male participants from the 2012/13 cohort. Three years after these individuals started on the pilot, the proportion who obtained a first NQF level 1 qualification increased by around 4ppts as a result of the programme. Boys from the 2014/15 cohort also experienced a positive impact (the proportion who achieved a first NQF level 1 qualification increased by almost 14ppts one year after starting on the pilot).
- By the end of the academic year in which female participants from the 2012/13 cohort started on the IF pilot, the programme had increased the proportion who achieved a first NQF level 1 qualification by around 7ppts. The IF also exerted a positive impact on girls from the 2013/14 cohort, having raised the proportion who achieved their first NQF level 1 qualification by 10ppts and 6ppts one and two years after they started participation, respectively.

The results of the tests which assessed whether, and to what extent, the impact of the IF pilot on the likelihood of achieving a first NQF level 1 qualification was higher for boys than girls are reported in Table 5.5. They suggested that:

- Within the 2012/13 cohort of participants, girls seem to have benefited more than boys from the IF pilot. The programme increased the proportion who achieved a first NQF level 1 by 8ppts more among girls than boys at the end of the IF start year. A similar finding was observed two years after the IF start year for the 2013/14 cohort (as indicated by a negative impact of -7.1ppts).
- On the other hand, results from the 2014/15 cohort suggested that the IF pilot had been more effective for boys than girls one year after they started on the programme, having increased the proportion who achieved their NQF level 1 qualification by over 12ppts more for males compared to females.

Table 5.5 Difference between impacts for male and female participants on the likelihood of achieving a first NQF level 1 qualification

	End of year	One year after	Two years after	Three years after
2012/13 cohort	-8.0**	-1.1	1.4	1.9
2013/14 cohort	2.5	-5.2	-7.1**	
2014/15 cohort	-0.5	12.3**		

** : Statistically significant at the 5 per cent level.

5.3.2 Impact on the likelihood of achieving a first NQF level 2 qualification

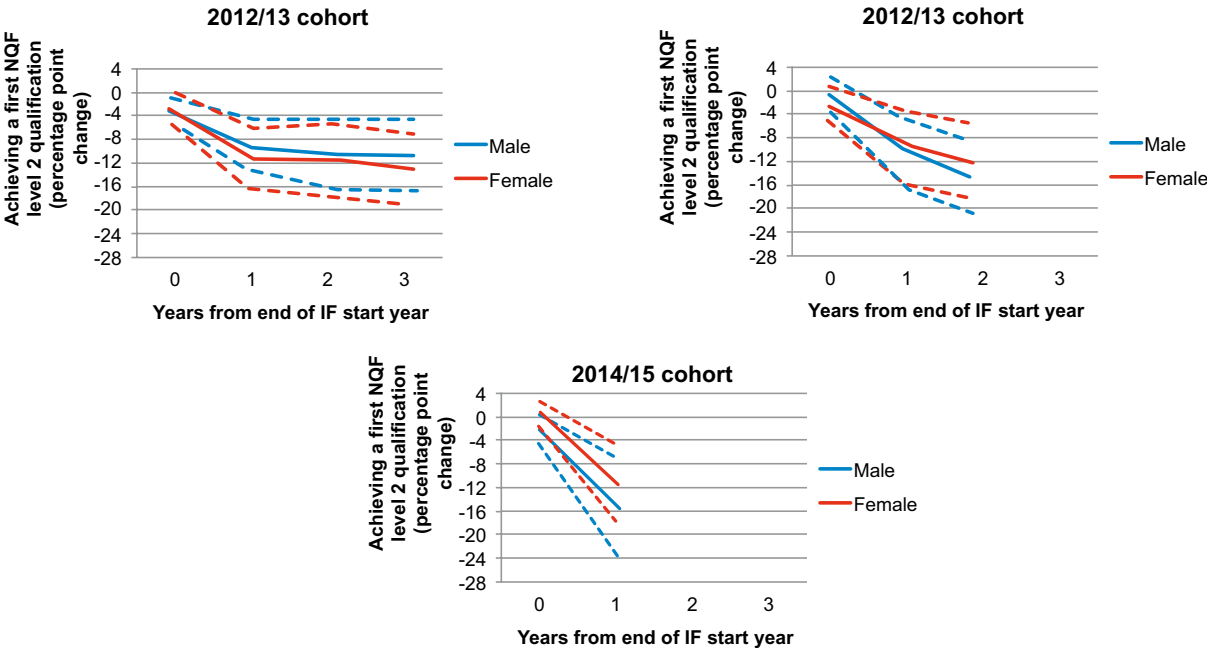
The findings from the by-gender analysis of the impact of the IF pilot on the likelihood of achieving a first NQF level 2 qualification, illustrated by means of Figure 5.5, indicated that:²⁵

²⁵The proportions of the participant and comparison group used to estimate impacts are reported in Table G.2 in Appendix G.

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

- The IF pilot reduced the likelihood of male participants from the 2012/13 cohort achieving a first NQF level 2 qualification. By the end of the academic year in which male participants from this cohort started on the programme, the proportion who obtained this qualification was reduced by 3ppts. Three years after, the estimated reduction was of around 11ppts).
- Negative impacts were also observed for male participants from the 2013/14 and 2014/15 cohorts. For example, two years after having started on IF pilot, the programme had reduced the proportion of boys from the 2013/14 cohort who achieved a first NQF level 2 qualification by 14ppts. Only one year after their IF start, boys from the 2014/15 cohort experienced a 16ppts reduction in the likelihood of achieving the same qualification.
- Female participants from all cohorts were found to have been negatively affected by the IF pilot. At the end of the follow-up periods observed (which varied depending on the cohort), the pilot had reduced the proportion who achieved a first NQF level 2 qualification by some 12ppts (the impact estimate was larger for girls from the 2012/13 cohort).

Figure 5.5 Impact of the IF pilot on the likelihood of achieving a first NQF level 2 qualification at each time point between the end of the IF start year and three years after, by gender of participant



The results of the tests which explored the existence of a difference between the IF impact on the likelihood of achieving a first NQF level 2 qualification for male and female participants are reported in Table 5.6. The negative impact difference (-2.9ppts) detected at the time point 'End of year' for the 2014/15 cohort indicates that, by the end of the IF start year, the IF pilot had been more effective for girls than boys from this cohort.

Table 5.6 Difference between impacts for male and female participants on the likelihood of achieving a first NQF level 2 qualification

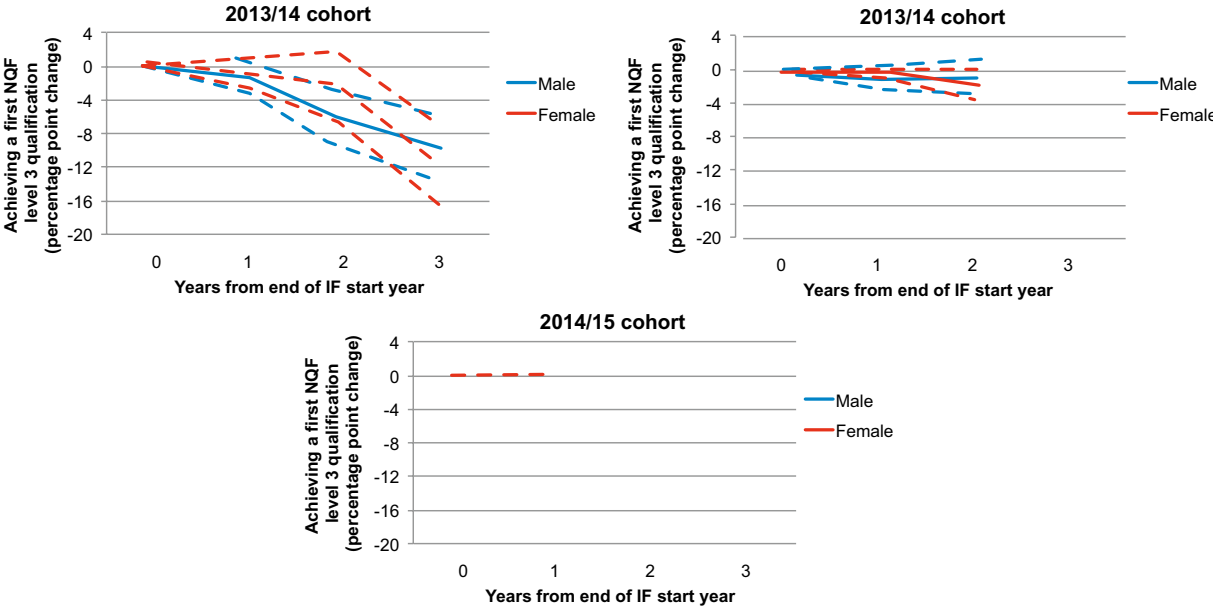
	End of year	One year after	Two years after	Three years after
2012/13 cohort	-0.4	2.0	1.0	2.6
2013/14 cohort	2.0	-0.4	-2.3	
2014/15 cohort	-2.9**	-3.8		

** : Statistically significant at the five per cent level.

5.3.3 Impact on the likelihood of achieving a first NQF level 3 qualification

The by-gender impact of the IF pilot on the likelihood of achieving a first NQF level 3 qualification is illustrated in Figure 5.6.²⁶

Figure 5.6 Impact of the IF pilot on the likelihood of achieving a first NQF level 3 qualification at each time point between the end of the IF start year and three years after, by gender of participant



The findings can be summarised as follows:

- Two and three years after male participants from the 2012/13 cohort started on the IF pilot, the programme had reduced the proportion who achieved their first NQF level 3 qualification by 6ppts and 9ppts, respectively.
- Three years after starting on the projects, the proportion of female participants from the same cohort who achieved this qualification was reduced by 12ppts as a result of the pilot.

²⁶The proportions of the participant and comparison group used to estimate impacts are reported in Table G.3 in Appendix G.

No statistically significant differences were found between the impact of the IF pilot on the likelihood of achieving a first NQF level 3 qualification for girls and boys.²⁷

5.3.4 Impact on school attendance and behaviour

The results of the by-gender analysis of the impact of the IF pilot on the likelihood of participants being persistently absent and excluded from school, respectively, are shown in Table 5. 7.²⁸

Table 5.7 Impact of the IF pilot on the likelihood of participants being persistently absent and excluded from school in the academic year after starting the pilot, by gender of participant

	2012/13 cohort	2013/14 cohort
Male		
Persistently absent (impact in ppts)	2.5	4.6**
Excluded (impact in ppts)	5.3	0.2
Female		
Persistently absent (impact in ppts)	-5.0	5.5**
Excluded (impact in ppts)	5.1	5.3**

Base (Male): 321 participants and 287 comparators (Persistently absent), and 352 participants and 316 comparators (Excluded) for the 2012/13 cohort; 238 participants and 231 comparators (Persistently absent), and 271 participants and 255 comparators (Excluded) for the 2013/14 cohort.

Base (Female): 343 participants and 325 comparators (Persistently absent), and 354 participants and 335 comparators (Excluded) for the 2012/13 cohort; 272 participants and 261 comparators (Persistently absent), and 276 participants and 267 comparators (Excluded) for the 2013/14 cohort.

** : Statistically significant at the five per cent level.

The results indicated that:

- The IF pilot increased the proportion of male participants from the 2013/14 cohort who were recorded as being persistently absent by 4.6ppts in the academic year after these individuals started on the pilot.
- A beneficial effect was observed for female participants from the 2012/13 cohort. These experienced a 5ppts reduction in the proportion recorded as persistently absent in the academic year after they started on the IF. However, among the same individuals, the pilot was found to have increased the proportion excluded from school by the same figure.
- The proportions of female participants from the 2013/14 cohort who were persistently absent and excluded from school were both found to have increased by over 5ppts as a result of the IF pilot.

²⁷See Table H4 in Appendix H.

²⁸The proportions of the participant and comparison group used to estimate impacts are reported in Table G.5 in Appendix G.

Finally, for what concerns differences in impact on behavioural outcomes, the IF pilot was found to have been more detrimental for boys than girls from the 2012/13 cohort in that findings indicated that the programme increased the proportion who were persistently absent by 7.4ppts more among the former than the latter (see Table 5.8).

Table 5.8 Difference between impacts for male and female participants on the likelihood of being persistently absent and excluded from school in the academic year after starting the pilot

	2012/13 cohort	2013/14 cohort
Persistently absent	7.4 **	-0.9
Excluded	0.2	-5.1

5.4 Discussion of the findings and further considerations

The results of the by-subgroup impact analyses are very difficult to summarise as estimated impacts varied depending on the cohort and outcome explored. Furthermore, virtually no evidence was available to interpret findings for specific projects, and little was known as to why impacts might have differed by round of commissioning or participant’s gender.

Overall, the results of the by-round subgroup impact analysis seem to suggest that the Round One projects were more effective than the Round Two projects in supporting participants to obtain their first NQF level 1 qualification. In addition, Round One projects were found to be less detrimental than Round Two projects in their attempt to help participants to achieve their first NQF level 2 qualification. These findings seem counterintuitive considering that all outcomes attracted higher payments in Round Two compared to Round One projects. One may indeed argue that higher payments translate into lower pressure on cash flow for Round Two projects compared to Round One projects, making the former projects more likely to achieve planned outcomes for target individuals than the latter projects (assuming that the cost of delivering the programme per participant was the same for both rounds of commissioning).²⁹

A possible explanation for these findings is that the availability of more financial resources per participant for Round Two projects compared to Round One projects resulted in the former projects providing more in-depth support to participants for achieving outcomes beyond those paid for by the DWP (and therefore not reflected in the outcomes considered by this evaluation). The negative by-round impact difference detected for participants from the 2012/13 cohort in relation to a first NQF level 3 outcome two years after starting on the projects (see Table 5.3) may also suggest that the Round Two projects were more successful in achieving level 3 qualifications only later in time because these qualifications were harder to achieve,

²⁹The qualitative evaluation of the IF pilot suggested that the increased scale of payments for Round Two projects made it ‘easier for these later projects to sustain cash flow. Partly as a result of this, Round Two projects were more likely to stick to their initial delivery plans and did not undergo the same degree of restructuring and reprofiling of outcomes, as did Round One projects’ (Thomas *et al.* 2016, page 40).

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

and helping participants required additional resources. However, these conclusions remain speculative. In addition, statistical tests indicated that high covariate imbalances existed in the comparison groups used to obtain impact estimates for Round 1 projects, and this compromised the validity of the inference made to explore the effectiveness of a round compared to the other.

Overall, the by-gender impact analysis detected higher impacts of the IF pilot on the likelihood of achieving first NQF level 1 and 2 qualifications for female than male participants. An exception was the 2014/15 cohort as, one year after starting on the pilot, boys experienced a higher increase than girls in their likelihood of achieving a first NQF level 1 qualification. However, the comparison groups used to estimate impacts were not balanced in terms of all individual characteristics considered for either subgroup (the imbalance was more pronounced for male participants), and the results of the by-gender analysis should be considered in light of this limitation.

6 SROI analysis

This chapter concerns analysis of the ‘Social Return on Investment’ (SROI) for the Innovation Fund (IF) pilot, and looks at the financial values achieved. It begins with simple measures of value achieved, based on contractual data, before bringing in results from the impact evaluation. The calculation of SROI is based, in part, on the calculation of impact so that the caveats applying to the impact evaluation also apply to the SROI analysis of financial value.

We have explained the reasons that the impact estimates are subject to uncertainty, and this caution must also apply to the SROI analysis reported in this chapter.

6.1 Background

6.1.1 Social Return on Investment (SROI)

The SROI approach has several similarities to the economic method known as ‘cost-benefit analysis’ (CBA), and in practice the differences may be quite nuanced. However, the CBA is generally employed to evaluate larger government projects, whilst SROI tends to be used to look at social enterprises and their wider impact (see Arvidson *et al.*, 2013).

This ‘social return on investment’ (SROI) helps to assess whether the benefits produced by projects, including the wider intangible benefits, exceeded the costs – that is the scale of return on investments. The SROI model outlines a methodology for calculating value as well as prescribing a set of principles for the evaluation of costs and benefits.

An ideal SROI investigation involves a range of stakeholders, and reviewing the inputs, outputs, outcomes and impacts made and experienced in relation to the projects. In practice, not all such data may be available to the analyst. For this project, the SROI analyst visited a number of the project sites and discussed the range and different kinds of benefits anticipated by them.

Some of the key problems experienced in conducting SROI evaluations are questions of deadweight and attribution. In this evaluation, a design that included a comparison group has been used to estimate the impact although there are important concerns about the matching of programme participants with that comparison group, which increases uncertainty about the results. Even so, it is helpful to provide a sensitivity analysis (one version of this appears in section 6.3.1) to indicate uncertainty about key elements of the SROI, particularly concerning deadweight and attribution, or more rarely the costs. The largely negative results from the impact evaluation imply that the costs of the programme are likely to have exceeded the benefits.

6.1.2 Social Impact Bonds (SIBs)

A distinctive feature of the IF was that it was implemented using Social Impact Bonds (SIB). As explained before, a SIB is a contract between an investor and the public sector, with the non-state investor providing the financial resources needed to implement the project and the state committing to re-pay them for improved social outcomes. A defining feature of SIBs is that repayment is contingent upon pre-agreed social outcomes being achieved (SIBs are comprised within the broad category of PbR models, also known as Pay for Success³⁰).

³⁰Other names include: “advance market commitments”, “rate cards” and “social impact insurance”.

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

For a recent (American) comprehensive overview of the SIBs environment, see Federal Reserve Bank of San Francisco & Nonprofit Finance Fund (2017) *What Matters: Investing in Results to Build Strong, Vibrant Communities*.

The data for this SROI analysis comes from several sources. Data from contracts and bids enable us to look at the costs compared to the state's Willingness to Pay (WTP) for those outcomes. To this can be added the number of relevant outcomes, which helps to identify where any value was created. Lastly, the impact evaluation enables analysis of the 'deadweight' of the programme, which is needed to calculate the ultimate value of the programme.

A key element of the SIB approach is the 'rate card', the set of prices paid in return for achieving a number of different outcomes specified in advance (see Table 6.1). This represents the WTP of the state body (DWP) against those key outcomes of interest. In principle, the WTP could be calculated in several different ways. First, as the direct cashable savings as a result of achieving outcomes which is the basis of the rate card – specifically, "These figures are based purely on the Department's benefit saving", (DWP, 2011: page 32), or expressed as "the associated future benefit savings" (DWP, 2011: p. 8). Alternatively, financial returns could have been determined based on the wider social benefits of attaining such outcomes; or, somewhere in between the social value and the direct cash savings value.

Table 6.1 Innovation Fund Rate Card for outcomes

	Round One	Round Two
Improved attitude to school/education	-	£700
Improved attendance at school	£1,300	£1,400
Improved behaviour at school	£800	£1,300
QCF Accredited entry level qualifications (below GCSE)	-	£900
First QCF Level 1 Qualification	£700	£1,100
First QCF Level 2 Qualification	£2,200	£3,000
First QCF Level 3 Qualification	£3,300	£5,100
Entry into Education at NQF Level 4	£2,000	-
Successful Completion of an ESOL Course	£1,200	-
Entry into First Employment	£2,600	£3,500
Entry into Sustained Employment	£1,000	£2,000

Source: DWP specification of the Innovation Fund.³¹

Investors/providers bid for the IF contract with the knowledge of the amounts payable for the different outcomes and the expectation that their own costs will be invoiced at lower rates. There are therefore two ways in which the SIB may produce added values compared to other or alternative arrangements. The first is ways in which the societal benefits of the programme exceed the WTP embedded within the 'rate card' which was based on direct cashable savings. The second is the difference

³¹For details see: <https://data.gov.uk/data/contracts-finder-archive/download/169814/de2c286f-7fd2-470d-8ca0-47accb9953dd> which includes the original Invitation to Tender.

between the bid prices from potential providers and the overall WTP as set out in the rate card, which might relate to the efficiency of the provider. The latter is the easier calculation to make, and that is where we start in section.

6.2 Costs and benefits of the IF

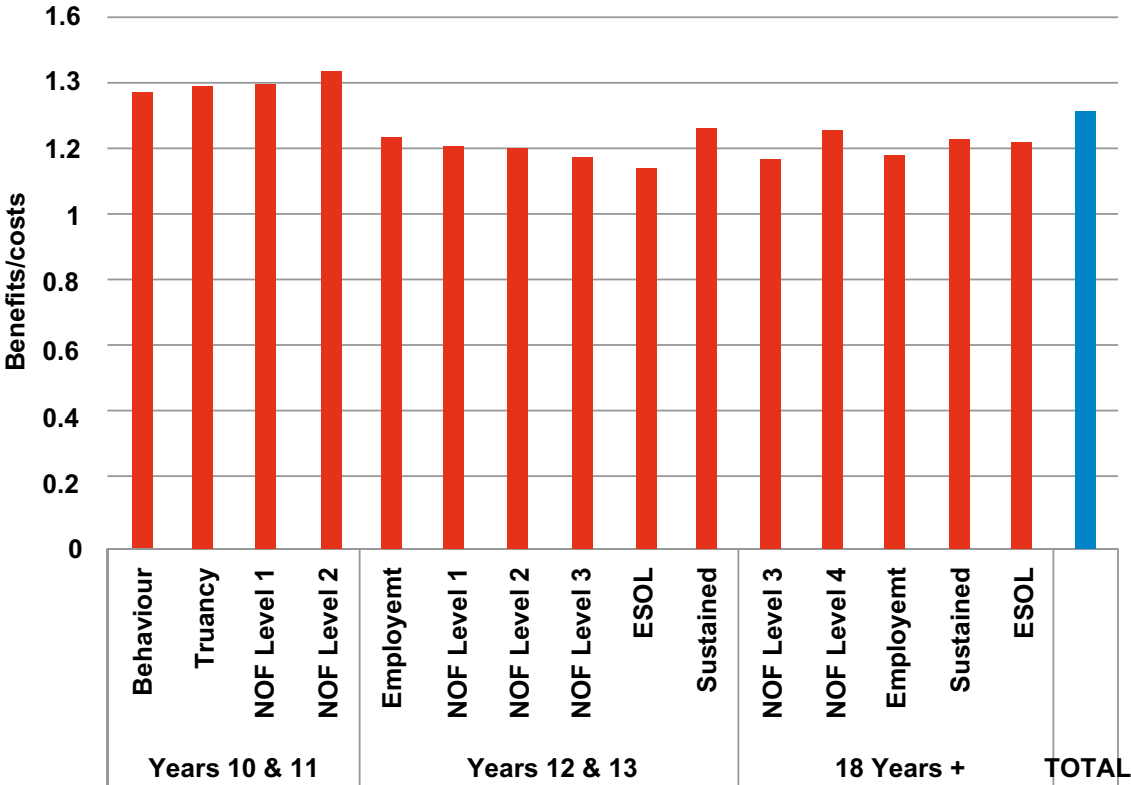
6.2.1 Direct costs and cashable benefits

Given that the benefits include a number related to education, there is already good evidence of the likely returns to undertaking education, and these returns are often strongly positive across a range of qualifications and context (see Psacharopoulos and Patrinos, 2004). There is also good evidence on the value of moving people off benefits and into paid work.

We start by treating costs as the invoiced cost of the projects (exclusive of VAT), and the benefits as relating to the state's WTP. The ratios of benefits to costs for Round One and Round Two projects are shown in Figure 6.1 Benefit-cost ratio for Round One outcomes and Figure 6.2 Benefit-cost ratio for Round Two outcomes, respectively. We would expect such ratios to be greater than 1 for there to be any savings at all from the projects.

In all cases, across the separate outcomes the ratio of benefits to costs exceeded 1, confirming that providers delivered all projects at a discount relative to DWP's WTP. The extent of discount varied, but across the board the average was 1.31 – the DWP was willing to pay 1.31 times as much as it actually paid. There was evidence of slightly better returns for the projects involving younger age groups.

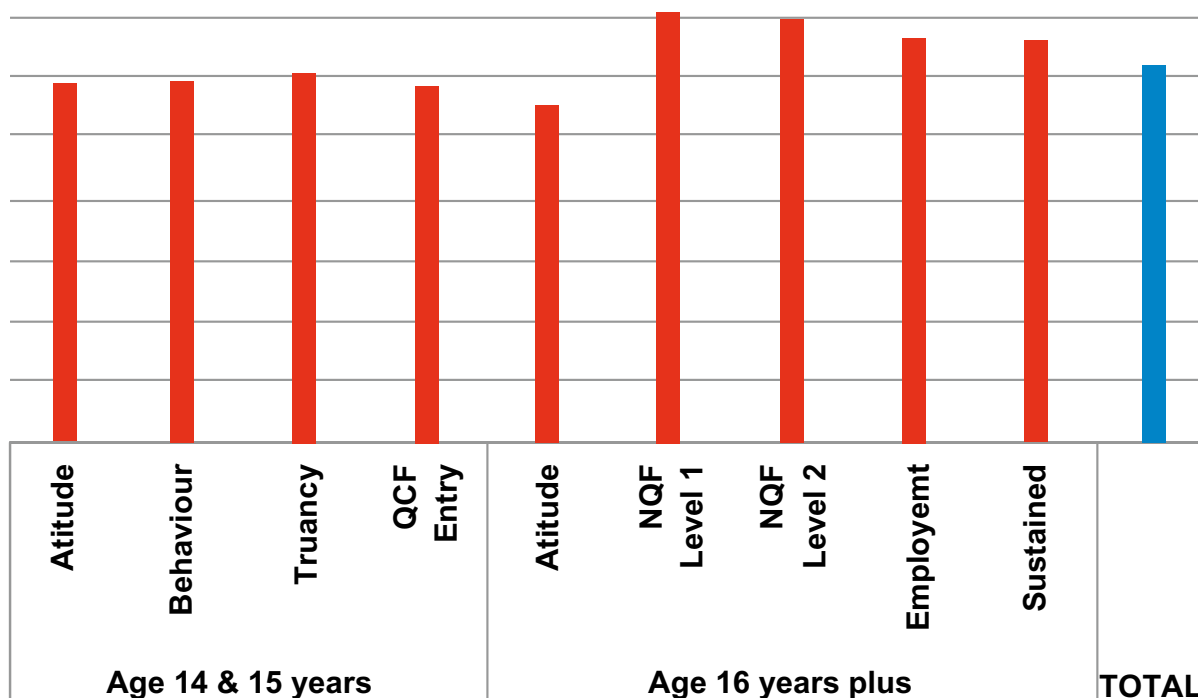
Figure 6.1 Benefit-cost ratio for Round One outcomes



Source: analysis of DWP accounting data.

The relevant rate card and bid prices were changed in Round Two of the IF programme. The ratio of WTP against actual payments fell a little to 1.25, perhaps reflecting a greater alignment of the costs and the rewards gained through learning within the projects and commissioners. There is a flatter profile, indicating limited differences in returns according to the different activities, which may be expected as a market matures.

Figure 6.2 Benefit-cost ratio for Round Two outcomes



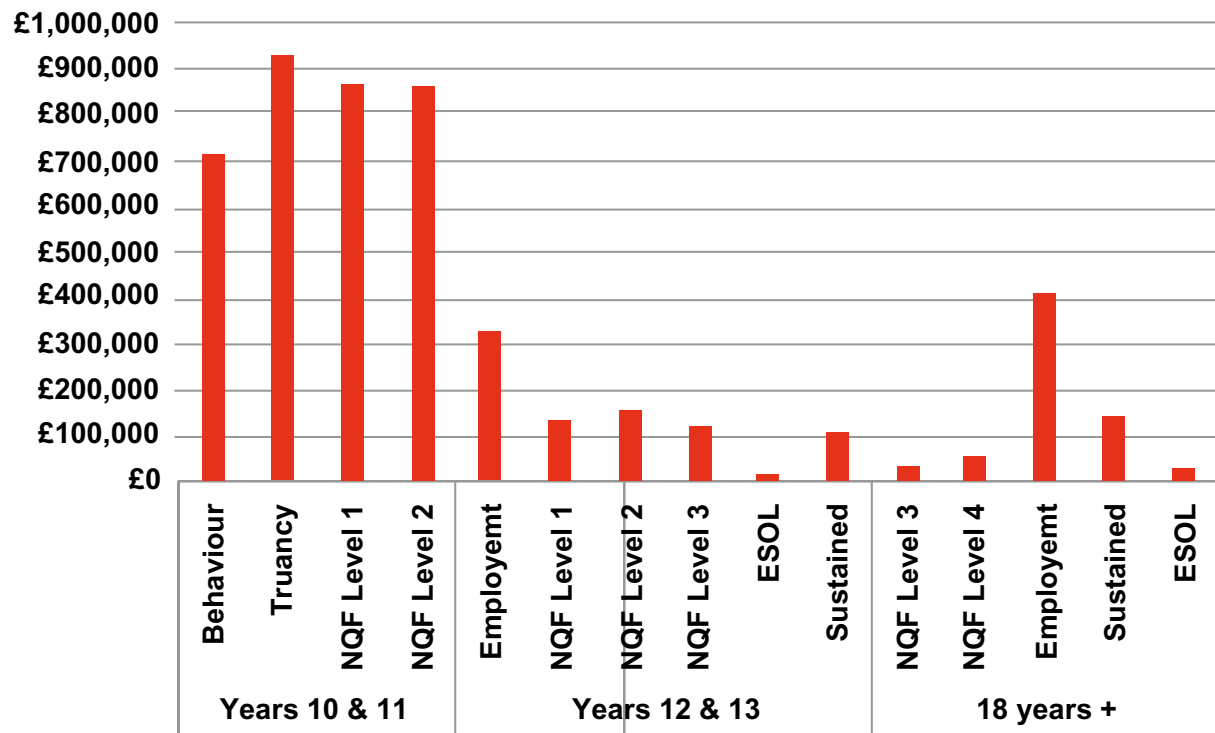
Source: analysis of DWP accounting data.

The benefit-cost ratios of 1.31 and 1.25 suggest a positive performance from the IF, but at a relatively modest level. Many SROI and cost-benefit studies give ratios that are many times this level. As examples, a systematic review in the area of public health based on 40 studies found that “SROI ratios varied across the different public health areas, with the highest ratio of 65.0:1 reported in a study in child health and the lowest ratio of 1.1:1 reported in a health promotion SROI study” (Banke-Thomas *et al.*, 2015: page 8). An evaluation of a transport to employment scheme in rural Scotland found an SROI ratio of 3:1 (Wright *et al.*, 2009). A study of ‘community befriending’ found an SROI ratio of 6.5:1 (Arvidson, Battye and Salisbury, 2014). Investment in two community schools was found to have SROI ratios that were greater than 10:1 (Martinez and Hayes, 2013).

The benefit-cost ratios show the value created in terms of benefits to costs. This is based on the ratio of the costs claimed by the providers for generating particular outcomes compared with DWP’s WTP for those outcomes. However, whilst such information is useful in signalling the added value of each outcome, it does not indicate where the total extra value created would have occurred. The total benefits also depend on the numbers of people achieving those outcomes; for example, the costs for an outcome might have been well below the benefit, but few such outcomes might have been delivered in practice. Hence, we need to look at the types of outcomes where any value was achieved.

As we show in Figure 6.3 for the first round, most of the overall benefits of the programme (differences between WTP and actual payments) came from, first, the younger age group and improved school performance (truancy and behaviour) and in educational attainment (NQF level 1 and 2 qualifications gained), and, to a lesser degree, from movements into paid work for the older end of the age range.

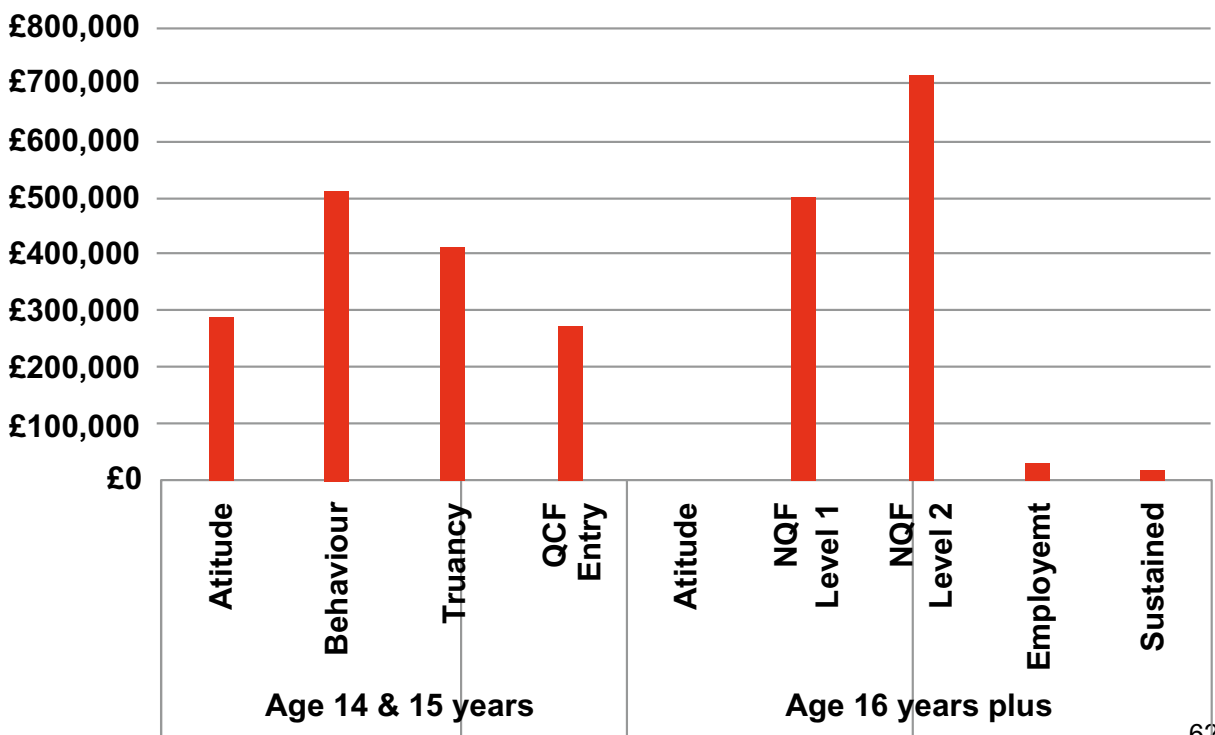
Figure 6.3 Activities contributing to benefits achieved: Round One



Source: analysis of DWP accounting data.

The balance of benefits attained was perhaps less unequal across age groups and activities at the second round (see Figure 6.4 Activities contributing to benefits achieved: Round Two). However, as with the first round, the outcomes relating to additional employment were much less common than the outcomes relating to (a) school attendance and (b) achieving NQF level 1 and level 2 qualifications.

Figure 6.4 Activities contributing to benefits achieved: Round Two

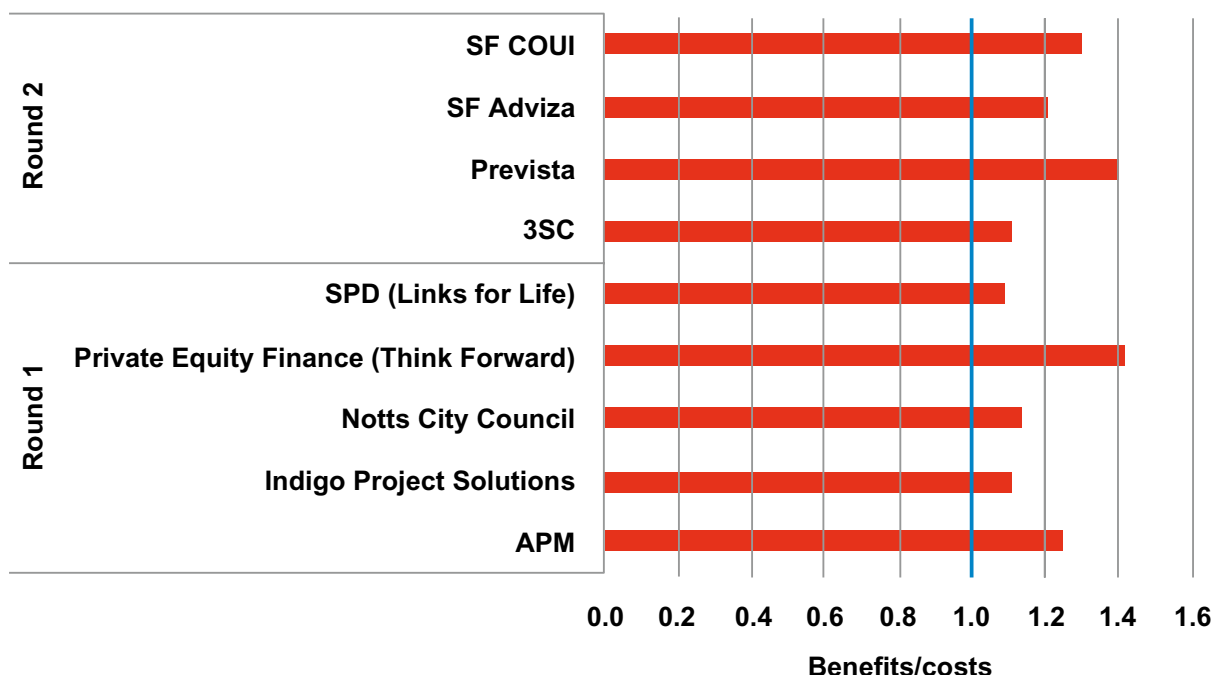


Source: analysis of DWP accounting data.

6.2.2 Provider-level analysis

In each round, there was a mix of different 'discounts' offered compared to the rate card. As shown in Figure 6.5 Benefit-cost ratios by contractor (averages across the different outcomes achieved in practice), the discounts varied from a high of 1.4 to a low of around 1.1. An alternative way of looking at this is that a benefit-cost ratio of 1.4 may be expressed as a discount of around 29 per cent compared to the rate card, while a ratio of 1.1 represents a discount of around 9 per cent.

Figure 6.5 Benefit-cost ratios by contractor (averages across the different outcomes achieved in practice)



Source: analysis of DWP accounting data.

6.2.3 Other approaches to monetising the IF benefits

SROI evaluations often produce very much higher benefit-cost ratios than those listed above – often closer to 5:1, but frequently more than 10:1. This is partly because they tend to include rather higher valuations for the different kinds of benefits achieved, going beyond the direct savings to benefits spending used above.

One approach, based on wellbeing, produces much higher estimates of benefits from changed outcomes. (See Trinder *et al.* (2013) for some recent estimates, and Fujiwara and Campbell (2011) for discussion of the underlying rationale and issues.) Recent estimates have included a suggested figure of £7,900 for making someone *more employable* (rather than the much lower figure on the DWP rate card of £2,600/£3,500 for moving into first employment). Analysis by Fujiwara suggests that avoiding depression and anxiety may be worth benefits amounting to £43,453 for a single person, whilst a move from unemployment to employment should be factored in with a value of £8,700 (Fujiwara, 2013), or sometimes more in later studies.

There are not well-established figures for each of the outcomes that were part of the evaluation, although as discussed there are existing benchmarks both for education and employment. If we were to substitute the larger values of evidenced wider benefits of achieving education, employment and wellbeing outcomes for those relating to the cashable benefits, then the SROI ratios would be 4-5 times higher than those shown above, and hence in the range of 5:1 or 6:1 – numbers rather more in line with estimates found for other programmes that have been evaluated using the SROI methodology. The above analysis looks at the results achieved without considering what might have happened in the absence of the IF pilot, and therefore how far the programme generated genuine additional value.

However, we do not pursue the larger estimates of value in too much detail because of the issues of deadweight found in the programme which tend to considerably offset the apparent benefits calculated. This is discussed below.

6.3 Calculating ‘additionality’

6.3.1 Deadweight and the impact estimates

The above calculations assume that all the positive outcomes recorded may be attributed to the programme. In other words, they imply that such outcomes would not have occurred in the absence of the intervention. For most policy interventions that is seldom the case – some people would have moved into work, or obtained a higher qualification, even if the programme had not been in place (a scenario that might be termed the counterfactual, or “business as usual”). This feature is known as ‘deadweight’. If 100 people gain from a programme, but 30 of them would have gained the same outcome in the absence of the programme, then there is 30% deadweight. A similar concept is that of ‘additionality’, the extent to which the outcomes are additional to those that would have anyway occurred. In some studies, there may also be concern about ‘displacement’, where others are negatively affected by an intervention (e.g. non-participants find it harder to get jobs in competition with the participants). This seems unlikely to be relevant for this study.

When conducting policy interventions, one means of ascertaining the level of deadweight is through a Randomised Controlled Trial (RCT). However, on grounds of ethics and logistics (and other reasons) these are not always feasible. In such circumstances, we rely on observational studies that seek to estimate the counterfactual, what would have happened if the IF had not been in place.

It should be clear, however, that the benefit-cost estimates above are quite vulnerable to even low levels of deadweight. If the ratio is only 1.25 (excluding the potential wider social benefits), as in the second round of the IF, then the programme would cease to yield any kind of positive return if the deadweight was as low as 20 per cent (other things being equal). In practice, the level of deadweight found in evaluations tends to be very much higher than that. This is likely to mean that the programme would not be found to be cost-effective, at least using the metric of WTP as the measure of benefit (then offset by the costs paid out).

As we have said above, there are reasons to be cautious about the key results of the impact evaluation, and this increases the uncertainties about evaluating the financial effects of the IF programme. As it stands, the impact evaluation implies that most of the scheme’s positive outcomes would have been achieved even if the scheme had not existed, and in some cases that the results from the statistical analysis suggests that participants in the programme achieved poorer outcomes than the comparison group. Moreover, the kinds of outcomes where the IF may have made a positive difference (e.g., lower level qualifications) do not provide particularly high rates of return either to society or to individuals.

Given the results of the impact study imply few positive outcomes, and relatively small additionality from even those positive outcomes, estimates of SROI and benefit-cost ratios in general will be below 1. This is, of course, dependent on the veracity of the impact evaluation based on Propensity Score Matching (PSM) (we have already exposed potential concerns about this approach).

6.3.2 Further caveats

It is also worth a reminder that there are other key issues concerning the evaluation of a particular programme, and how that may be generalised to a more 'scaled up' version. With new programmes, like the IF, there are difficulties in extrapolating from the original scheme what would be the case if such programmes were significantly increased in scale. Some of these considerations might suggest a better effect, whilst others argue that effects would be negative. For instance, providers may want to be seen as strong bidders for potential work, which might provide an incentive to submit more competitive bid prices. On the other hand, there is also greater uncertainty about what is practicable, which would lead providers to act in more cautious ways in making bids. Providers may also be concerned that too low a level of apparent 'discount' on the WTP would not show the project in a strong light, encouraging more discounting. A further issue is that pilot projects are often set up with considerable enthusiasm and with the most dedicated staff. When a project is rolled out, and there is a need for provision where perhaps it did not previously exist, the commitment levels may be harder to sustain over time. Conversely, there may be key points that are learned during the initial phase of the programme that identify better approaches.

These considerations do not point us towards particular conclusions about whether a project delivered at a larger scale would be either more or less cost-effective than the early stage projects. However, they do raise issues about generalising from the smaller pilot-scale project to a bigger scheme.

7 Conclusions

The Innovation Fund (IF) pilot was a £30million initiative delivered in the UK between April 2012 and November 2015 to support young people aged between 14 and 24 who were disadvantaged or at risk of being disadvantaged. The programme comprised ten projects which used a 100 per cent Payment by Results (PbR) SIB funding model.

This report illustrated the findings from the impact evaluation of the IF pilot, alongside the results of the Social Return on Investment (SROI) analysis. The impact evaluation explored the extent to which the IF pilot was successful in helping programme participants to engage in a range of activities which were likely to improve their chances of being employed and reduce their longer-term dependency on benefits. The SROI analysis assessed the ability of the pilot to generate benefits, including social benefits and fiscal savings, over and above the costs of the programme

7.1 Impact evaluation

The impact evaluation of the IF included two separate assessments, both covering young people who participated in pilot projects delivered in England. The first assessment used information collected through a survey of IF (Round One) participants, all aged 14 to 18 at the start of the programme, and a selected sample of non-participants similar to participants in several important respects.

The assessment explored the impact of the pilot on the likelihood of participants being in EET around one year after starting on the projects. The second assessment used administrative data from the NPD and ILR to investigate the impact of the IF pilot on participants in (Round One or Round Two) projects from three cohorts (namely, young people who started the IF, when aged 14 or 15, between November and April of the academic years 2012/13, 2013/14 and 2014/15). For each cohort, the analysis explored the impact of the pilot on the likelihood of achieving any of a first NQF level 1, 2 or 3 qualification. For the 2012/13 and 2013/14 cohorts, the impact of the programme on the likelihood of participants being persistently absent and the likelihood of them being excluded from school were also investigated.

The findings from the **survey-based impact assessment** suggested that:

- One year after starting the IF pilot, the proportion of participants who were in school or college and the proportion in a paid job were both reduced, suggesting a negative impact of the pilot (more young people would have been in education or employment had they not participated in the IF). On the contrary, the proportion in training was found to have increased, suggesting a positive impact (i.e., the proportion of participants in training increased as a result of the programme).
- Young people could have been involved in multiple activities at the time of the survey (e.g., they could have been in employment but also doing a traineeship). When a more aggregated outcome measure was used, the IF pilot was found to have reduced the likelihood of participants being in employment, education or training. This negative impact was greater for older IF participants (those aged 16 or over) than younger ones (14 and 15 year olds).

However, the above results should be viewed with caution. The observed negative impacts may not constitute genuine findings of a detrimental effect of the pilot. It is possible that the lack of suitable data capturing relevant personal characteristics and

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

circumstances associated with the individuals' NEET status (or the risk of becoming NEET, for younger participants) prior to being involved in the IF pilot (e.g., being a teenage parent, drug or alcohol misuser, young carer or gang member, and personal motivation) have resulted in comparison groups of non-participants intrinsically different from IF participants being used in the estimation of impacts. For example, if the IF participants were more disadvantaged (and therefore more difficult to support) than their comparators, this might have biased impact estimates downwards, underestimating the causal effect of the IF pilot on the likelihood of participants to achieve all outcomes considered. IF participants and the comparison group of non-participants could have also differed in terms of the support received under other programmes similar to the IF pilot (e.g., the Youth Engagement Fund and Youth Employment Initiative) but this could not be explored due to data limitations.

Similar to the survey-based analysis, the findings of the **impact assessment which used administrative data** could have been influenced by the inability of the analytical approach, due to data limitations, to control for all the predictors of programme participation and of the outcomes of interest. Hence, some caution is needed when interpreting the findings, which were as follows:

- The pilot was beneficial to participants in that it helped them achieve a first NQF level 1 qualification, mostly one year after they started on the programme. This finding is in line with the qualitative evidence of the pilot (Thomas *et al.*, 2016), which suggested that gains for participants were mainly in the form of lower-level qualifications (easier to achieve and delivered in-house by projects) and improvement in school attendance and behaviour.
- Adverse effects were detected with regard to the achievement of higher-level qualifications: the IF pilot was found to have reduced the proportion of participants from all cohorts who achieved their first NQF level 2 and level 3 qualifications (the largest negative impacts were observed at the latest point in time for each cohort).
- In addition to the possible limitations of the analytical approach (exposed above), the observed negative impacts could also reflect the incapacity of some projects to deliver staged outcomes. The IF qualitative evaluation (Thomas *et al.* 2016) documented that, due to cash flow constraints, several projects turned away from the staged-progression delivery model to pursue a 'rolling intake' model, based on large numbers of recruits and fewer outcomes per person achievable in a short time frame (notably, NQF Level 1 qualifications and improvements in school behaviour and attendance). Therefore, supporting participants towards the achievement of qualifications higher than NQF level 1 seems to have been very limited in reality, and some projects may have 'traded' the most ambitious individual outcomes (achievement of higher-level qualifications) with softer ones (improvement in attendance and attitude towards school among programme participants).
- Positive impacts on educational outcomes were higher and negative impacts more contained for later compared to earlier cohorts, possibly reflecting a learning process which benefited those who were supported after the initial implementation phase.
- This study found that the pilot increased the likelihood of being persistently absent and of being excluded from school in the year after having started on the IF. However, the measures used to capture these outcomes (not well-suited to fit the definition of outcomes paid for by the programme) are likely to have largely underestimated the benefits of the pilot and should therefore be considered cautiously.

An additional caveat, concerning the comparison groups used to estimate impacts, applies to the subgroup analyses. Non-participant comparators used to estimate the subgroup impacts for male participants and participants in Round One projects showed imbalances with respect to certain observed variables. That is, they were markedly different from participants in some important regards. This casts doubts on the validity of the statistical tests aimed at exploring whether the impact was more effective for one subgroup compared to the other (Round One versus Round Two, and boys versus girls).

7.2 SROI analysis

On the basis of willingness to pay as a metric of benefit, itself based on savings to benefits spending, the IF projects achieved value for money (the ratio of benefits to cost was around 1.3 for Round One projects and 1.25 for Round Two projects). More common SROI ratios based on wider social benefits (notably those derived from the 'well-being valuation' approach), would suggest SROI ratios about 4-5 times higher. However, the benefit-costs ratios computed, whatever the method of calculation, need to be tempered with the results of the impact evaluation. The impact results imply that most of the scheme's positive outcomes would have been achieved even if the scheme had not existed, and in some cases that programme participants achieved poorer outcomes than the comparison group. Based on these considerations, the real SROI ratios would be below one, suggesting that the IF did not achieve value for money overall or at an aggregate level. There was, however, evidence of a positive effect on level 1 educational outcomes.

7.3 Overall conclusions

The findings of this evaluation indicated that the IF pilot was beneficial to participants, in that it was found to have supported them to achieve NQF level 1 qualifications. However, there was also evidence suggesting that the proportion of participants achieving higher level qualifications was reduced as a result of the programme. It is possible that data constraints affected the analytical approach used to estimate impacts, and therefore the findings should be interpreted with caution. This caveat also restricts the ability of the SROI analysis to confidently state that the benefits of the IF exceeded its costs.

At the same time, it is important to note that the IF pilot was a new programme, designed to address the needs of a particularly disadvantaged group of young people, and therefore difficult to support. It is not surprising then that its delivery encountered a number of challenges, as documented in the qualitative impact report (Thomas *et al.*, 2016). It is possible that financial considerations have led some projects to 'trade' hard to achieve outcomes such as higher-level qualifications with softer outcomes (improvements in attitudes and school attendance). In the light of this, DWP may want to explore reconsidering the incentives within SIB models that seek to help the most vulnerable young people.

References

- Ainsworth, P. and Marlow, S. (2011). *Early Impacts of the European Social Fund 2007-13*. DWP In-House Research No. 3.
- Arvidson, M., Battye, F. and Salisbury, D. (2014). The social return on investment in community befriending. *International Journal of Public Sector Management*, 27(3), 225-240.
- Arvidson, M., Lyon, F., McKay, S. and Moro, D. (2010). *The ambitions and challenges of SROI*. University of Birmingham, Third Sector Research Centre Working Paper 49.
- Arvidson, M., Lyon, F., McKay, S. and Moro, D. (2013). Valuing the social? The nature and controversies of measuring social return on investment (SROI). *Voluntary Sector Review*, 4(1), 3-18.
- Banke-Thomas, A., Madaj, B., Charles, A. and van den Broek, N. (2015). Social Return on Investment (SROI) methodology to account for value for money of public health interventions: a systematic review. *BMC Public Health*, 15:582, pp 1-14. DOI 10.1186/s12889-015-1935-7.
- Barker, D., Cassidy, J. and Chen, W. (2017). Paying For Results. Reforming How The Public Sector Funds Social Services in Federal Reserve Bank of San Francisco & Nonprofit Finance Fund (2017), 68-78.
- Blundell, R. and Costa Dias, M. (2000). *Evaluation Methods for Non-Experimental Data*, Fiscal Studies, 21(4), 427-68.
- DWP (2011) *The Innovation Fund. Invitation to Tender. Specification and Supporting Information*. London: DWP.
- Federal Reserve Bank of San Francisco & Nonprofit Finance Fund (2017) *What Matters: Investing In Results To Build Strong, Vibrant Communities*. (Editors: Bartlett, V., Bugg-Levine, A., Erickson, D., Galloway, I., Genser, J. and Talansky, J.).
- Fujiwara, D. and Campbell, R. (2011). *Valuation techniques for social cost-benefit analysis: stated preference, revealed preference and subjective well-being approaches: a discussion of the current issues*. London: HM Treasury.
- Fujiwara, D. (2013). *The Social Impact of Housing Providers*. London: HACT.
- Martinez, L. and Hayes, C. (2013). *Measuring Social Return on Investment for Community Schools: A Case Study*. Children's Aid Society.
- Orr, L., Bell, S. and Lam K. (2007). *Long term impacts of the New Deal for Disabled People*, DWP Research Report No. 432.
- Psacharopoulos, G. and Patrinos, H. (2004). Returns to investment in education: a further update. *Education Economics*, 12(2), 111-134.
- Rosenbaum, P. R. and Rubin, D. B. (1983). *The Central Role of the Propensity Score in Observational Studies for Causal Effects*. *Biometrika*, 70, 41-55.
- Thomas, A. and Griffiths, R. (2014). *Innovation Fund pilots qualitative evaluation: Early implementation findings*. DWP Research Report No. 880.
- Thomas, A., Griffiths, R. and Pemberton, A. (2016). *Qualitative evaluation of the DWP Innovation Fund: Final report*. DWP Research Report No. 922.

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

Trotter, L., Vine, J., Leach, M. and Fujiwara, D. (2014). *Measuring the Social Impact of Community Investment: A Guide to using the Wellbeing Valuation Approach*. London: HACT.

Wright, S., Nelson, J., Cooper, J. and Murphy, S. (2009). An evaluation of the transport to employment (T2E) scheme in Highland Scotland using social return on investment (SROI). *Journal of Transport Geography*, 17(6), 457-467.

A Area-level matching

Table A.1 Matched areas

Project Name	Pilot LA	Matched LA
Triodos	Halton	Stockton-on-Tees
Triodos	Knowsley	Sunderland
Triodos	Liverpool	Newcastle upon Tyne
Triodos	Sefton	Kirklees
Triodos	St. Helens	Rotherham
Triodos	Wirral	Gateshead
Nottingham City Council	Nottingham	Coventry
Advanced Personnel Management	Birmingham	Sandwell
Advanced Personnel Management	Solihull	Dartford
Links to success	Newham	Greenwich
Links to success	Waltham Forest	Lambeth
Private Equity Foundation	Hackney	Enfield
Private Equity Foundation	Islington	Kensington & Chelsea
Private Equity Foundation	Tower Hamlets	Redbridge

B Generating pseudo start dates for non-participants

The educational and behavioural outcomes of participants were measured at the end of the academic year in which they started the IF pilot, and at the end of each of the next three academic years. However, those in the comparison group (non-participants) did not have a natural start date following which outcomes could be measured. Therefore, fictional (or 'pseudo') start dates were assigned to each non-participant so that a time-based comparison between the outcomes of participants and non-participants could be made. The technique employed to generate pseudo start dates aimed to align participants and non-participants with respect to calendar time and age.

The steps implemented to generate pseudo start dates for non-participants are as follows:

1. From the MI records of IF participants all the possible combinations of month and year of birth were identified. Separate datasets were then created for participants born in July 1987, August 1987, September 1987 and so on up to the latest month-year combination observed in the data (May 2001);
2. For each of these datasets, the proportion of participants who started the IF pilot in each possible month and year was observed.
3. The process adopted in Step 1 was replicated for all non-participant records, which were separated according to individuals' month and year of birth (this information was sourced from the NPD).
4. Within every dataset obtained at Step 3, a pseudo IF start month and year was randomly assigned to each non-participant so that the distribution of start months and years resembled the distribution observed for the participant dataset with the equivalent month and year of birth.

C Profile of IF participants

This section profiles the samples of IF participants used for the survey- and administrative data-based impact analyses, respectively, in relation to their baseline (pre-treatment) characteristics. When the participant sample used for the impact analysis based on administrative data is profiled, these characteristics (alongside with outcomes) are also shown for the sample of non-participants. Comparing participants and non-participants with respect to baseline characteristics aims to highlight differences between the two groups which can potentially explain the reasons that some individuals participated in the pilots and, at the same time, variation in the outcomes explored.

C.1 Impact analysis based on survey data

This section describes the features of IF participants who responded to the first wave of the IF survey. These individuals were profiled with respect to their main baseline demographic, educational, and family- and parent-related characteristics. Their distribution across pilot projects is also shown.

The information used to profile IF participants was mainly sourced from the responses to the survey, but MI or NPD variables were also used. Individuals' characteristics obtained from the NPD are observed prior to participation in the IF. In those instances where an IF participant's characteristic is observed subsequently to their start in the programme (e.g., because it was collected through the survey, after young people started the IF pilots), this characteristic is used to profile individuals only if it is either time-invariant or if it is plausible that it was not affected by the pilot. This is to avoid the possibility that the characteristics observed are the result of programme participation rather than its determinants. Statistics are based on the 891 IF participants who were included in the survey-based impact estimation.

C.1.1 Basic demographics

Table C.1 describes the main demographic characteristics of young people who were selected for participation in the IF. These characteristics include participants' age at IF start, gender, ethnic group, whether they were born in the UK and whether English was their main language. Participants' age at IF start was calculated using information about young people's date of birth (available from the survey) and programme start date (from the MI).

As Table C.1 shows, most of the programme participants were between 14 and 15 years old when they started on the IF as those within this age band make up around 55 per cent of the participant sample. The proportion of participants decreased with age: around one third of IF participants (32.3 per cent) were 14 years old, which compares to over one ninth (11.8 per cent) who were 18 years of age. IF participants were more likely to be male (63.5 per cent) and of white ethnicity (66.8 per cent). Around nine tenths of them (89.9 per cent) were born in the UK and over three quarters (76.1 per cent) had English as their main language.

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

Table C.1 Basic demographics of IF participants

Characteristic	
Age at IF start (median, in years)	15
Age at IF start is 14 (%)	32.3
Age at IF start is 15 (%)	22.6
Age at IF start is 16 (%)	18.6
Age at IF start is 17 (%)	14.7
Age at IF start is 18 (%)	11.8
Male (%)	63.5
White ethnicity (%)	66.8
Asian ethnicity (%)	14.8
Black ethnicity (%)	9.8
Mixed ethnicity (%)	6.7
Other ethnicity (%)	1.9
Born in the UK (%)	89.9
English is the first language (%)	76.1
<i>Base: 891 respondents.</i>	

C.1.2 Educational characteristics

IF participants are also profiled with respect to some of their educational characteristics, namely, their eligibility for Free School Meals (FSM) and the degree to which they have a Special Education Need (SEN). FSM eligibility and SEN status are among the most commonly used indicators of material disadvantage for young people. Information about these characteristics was sourced from the NPD.

Table C.2 Educational characteristics of IF participants

Characteristic	%
Eligible for FSM	42.9
No SEN	46.4
SEN School Action	21.9
SEN School Action Plus	23.8
SEN Statement	8.0
<i>Base: 891 respondents.</i>	

As Table C.2 shows, a very large proportion of IF participants (42.9 per cent) were eligible for FSM. This suggests that a substantial proportion of the young people who took part in the IF pilots were from low income families. Table C.2 also indicates that the majority of programme participants (53.6 per cent) were young people with a SEN. Almost one in four IF participants (21.9 per cent) had a SEN School Action, meaning that their school provided additional help to them in order to address their needs. A similar proportion (23.8 per cent) had a SEN School Action Plus, reflecting that the school had requested external help, such as from an educational psychologist or therapist, to support the young person (School Action Plus implies a more serious SEN than School Action). Eight per cent of the participant sample had a SEN Statement, which is a statutory assessment by the local authority conducted with children who have the most severe needs.

¹ Unfortunately, the NPD variable indicating whether the young person had been permanently excluded from school during the academic year could not be used in the analysis as information was missing for 89 per cent of the participant sample. Similarly, the variable denoting whether the young person was a persistent absentee had to be discarded due to the high level of missing observations.

C.1.3 Family- and parent-related characteristics

Survey participants were asked whether either of their parents stayed in education after the age of 16 and how many siblings they had. This made it possible to explore the composition of the IF participant sample with respect to these family- and parent-related characteristics.²

As reported in Table C.3, almost 80 per cent of IF participants reported that neither of their parents stayed in education after the age of 16. Only a small minority of programme participants (8.3 per cent) had no siblings while most of IF participants reported having more than one sibling.

Table C.3 Family- and parent-related characteristics of IF participants

Characteristic	%
Either parents stayed in education after the age of 16	22.1
No siblings	8.3
One sibling	25.6
Two siblings	24.0
Three siblings	17.5
Four or more siblings	24.6
<i>Base: 891 respondents.</i>	

C.1.4 Distribution across pilot projects

Table C.4 shows how IF participants who responded to the survey were distributed across the Round One pilot projects. The Indigo Project Solutions project was not included in the survey because, initially, sample sizes were expected to be small. Furthermore, the methodology for matching and contacting the comparison sample using England’s NPD was not feasible with education records in Scotland. These considerations meant that the IF survey covered the five projects in England.

As the table shows, over one third (35.5 per cent) of participants started the Triodos project, and around a quarter (25.6 per cent) were supported by Nottingham City Council. Around one in every five young people (19.6 per cent) among those in the participant sample was supported by APM Ltd, while Links to Success supported only one in every 25 (3.8 per cent). The remaining 15.5% of the IF participants were involved in the PEF project.

Table C.4 Distribution of IF participants across pilot projects

Pilot project	%
Triodos	35.5
Nottinghamshire City Council	25.6
APM Ltd	19.6
Links to Success	3.8
PEF	15.5
<i>Base: 891 respondents.</i>	

² Unfortunately, other parent/family variables initially considered for inclusion in the analysis could not be used. For example, a variable denoting whether either of the IF participants’ parents attained a university degree was discarded as the information was missing for 72 per cent of the participant sample.

C.2 Impact analysis based on administrative data

This section profiles the sample of IF participants used for the administrative data-based impact analysis with respect to their outcomes and baseline characteristics (both sourced from administrative data). This sample is also compared with the non-participant sample across the same features. The descriptive statistics reported in this section are calculated for the pooled sample of participants included in the three cohorts used to estimate impacts (2,817 individuals, aged 14 or 15 at IF start). The sample of non-participants (879,176 individuals aged) was composed of all young people from English areas not covered by the pilots and aged 14 or 15 at their pseudo-start date.

C.2.1 Outcomes

The impact evaluation based on administrative data used five outcomes which attempted to capture some of the participants' outcomes upon achievement of which contractors could claim payments from the DWP.

These outcomes can be divided into two sets:

- **Educational outcomes** – These denoted whether young people achieved each of three school attainments among a first National Qualifications Framework (NQF) level 1, level 2 and level 3 qualification.
- **Attendance/behaviour-related outcomes** – These included two variables: the first indicated whether the pupil missed 15 per cent or more of their school sessions in a given academic year, and was meant to denote whether they had been persistently absent from school; the second indicated whether the pupil had been excluded from school during the academic year in question (both fixed and permanent exclusions were considered).

Table C.5 profiles the sample of IF participants and non-participants with respect to these outcomes. For the purpose of profiling individuals, educational outcomes were observed only at the end of the year in which young people started on the programme. Instead, attendance/behaviour-related outcomes were observed over the academic year following the academic year in which individuals started the pilot (for non-participants, the academic year in which their pseudo start date fell).

³ The complete list of outcomes is provided in the IF guidance documents for Round 1 and Round 2, available at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/626291/round-one-provider-guidance.pdf and https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/626292/round-two-provider-guidance.pdf

⁴ Unfortunately, the impact of the IF pilot on the 'QCF Entry-level' outcome could not be explored. This was because information about pupils' QCF Entry-level qualifications was available only up to (and including) the academic year 2012/13, and this meant that it was not possible to estimate impacts on this specific educational outcome for participants from the 2013/14 and 2014/15 cohorts at any point in time following their start on the IF pilot. Neither was this possible for participants from the 2012/13 cohort one, two and three years after the academic year in which they started on the IF pilot. Furthermore, note that this educational outcome was relevant only to the IF Round Two projects.

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

Table C.5 Descriptive statistics on outcome variables (all cohorts)

	Participants (%)	Non-participants (%)
First NQF level 1	66.7	67.6
First NQF level 2	7.8	25.1
First NQF level 3	0.0	0.0
Persistently absent	6.1	1.5
Excluded	9.4	5.3

Base: 2,817 participants and 879,176 non-participants (First NQF level 1, 2 and 3); 1,697 participants and 431,137 non-participants (Persistently absent); and 1,792 participants and 439,580 non-participants (Excluded).

As indicated in Table C.5, by the end of the academic year in which participants started the programme, 66.7 per cent of them had achieved a first NQF level 1 qualification. A slightly higher proportion (67.6 per cent) was observed to have obtained the same qualification among non-participants by the end of the year in which their pseudo start date fell. The achievement of a first NQF level 2 qualification was much more frequent among non-participants (25.1 per cent) than participants (7.8 per cent). No participant achieved a first NQF level 3 qualification by the end of the year in which they started the IF pilot, and a very small proportion of non-participants (rounded to zero in the table) obtained such qualification. In the year after the IF start (or pseudo-start) year, a larger proportion was found to have been persistently absent from school among participants (6.1 per cent) compared to non-participants (1.5 per cent), and participants were twice as likely as non-participants to have received an exclusion (9.4 versus 5.3 per cent).

While the statistics presented in Table C.5 suggest that non-participants outperformed participants across all outcome measures considered, this does not necessarily reflect a negative effect from participating in the IF pilot. The crucial issue in attributing a causal effect (impact) to the programme lies in understanding whether non-participants are comparable to participants across the spectrum of personal characteristics observed.

C.2.2 Baseline characteristics

Table C.6 shows the baseline characteristics of IF participants and non-participants.

Table C.6 Descriptive statistics on baseline variables (all cohorts)

	Participants	Non-participants
Age at IF start		
14 (%)	57.2 ***	64.8
15 (%)	42.8 ***	35.2
Gender		
Female	50.2 ***	48.9
Male	49.8 ***	51.1
Ethnicity		
White (%)	77.8 ***	84.4
Black (%)	7.3 ***	3.9
Asian (%)	7.6	7.1
Mixed (%)	5.6 ***	3.8
Other (%)	1.7 ***	0.9
English as a first language (%)	93.6	93.0
Special Educational Needs (%)	33.2 ***	13.1
Eligible for Free School Meals (%)	43.2 ***	15.7
KS2 English attainment		
Below expected level (%)	40.1 ***	17.4
At expected level (%)	49.5	49.6
Above expected level (%)	10.4 ***	33.0
KS2 Maths attainment		
Below expected level (%)	30.8 ***	14.0
At expected level (%)	58.4 ***	52.8
Above expected level (%)	10.8 ***	33.3
Persistently absent (%)	5.4 ***	1.5
One or more exclusions (%)	15.7 ***	4.7

The two groups had markedly different characteristics. To start with, IF participants were less likely to be younger than non-participants as a smaller proportion of individuals (57.2 versus 64.8 per cent) were aged 14. They were also more likely to be of a non-white ethnicity and be more disadvantaged in several regards. For example, one-third (33.2 per cent) of IF participants had a SEN, which compared to only one-eighth (13.1 per cent) among non-participants. IF participants were disproportionately more likely to be eligible for FSM than non-participants (43.2 versus 15.7 per cent), and had larger proportions who scored below the expected level in both English and Maths. Importantly, the proportions of young people who were recorded as being persistently absent and excluded from school were around three times as likely among IF participants compared to non-participants. Pilot participants attended schools located in more deprived areas than non-participants, as suggested by a comparison between the IMD scores of the two groups.

Statistical tests confirmed the existence of differences between IF participants and non-participants with respect to all the aforementioned characteristics. An important implication of this finding is that two groups of young people cannot be considered similar to each other, and therefore the outcomes observed among non-participants cannot plausibly provide a proxy for the counterfactual (the outcome that IF participants would have experienced had they not started the pilot). This suggests that the PSM approach can be implemented to identify, among all non-participants, a subset who resembles IF participants reasonably well, and whose outcome can be used in the estimation of impacts.

C.2.3 Distribution of IF participants across pilot projects

An important question that this evaluation sought to answer is whether the impact of the IF pilot varied across the seven projects explored. However, the extent to which this issue could be investigated was limited by the sample size observed for some projects. Table C.7 shows the number of participants in each project, separately for each cohort of starters. In the table, square brackets denote small sample sizes (fewer than 50 observations) or figures suppressed to avoid using them to retrieve small sample sizes in other cells. Those projects which included too few individuals over a particular time period (cohort) were excluded from impact assessment. Impact estimates were produced only for three projects (Energise, Teens & Toddlers and Triodos). However, the results are not illustrated in this report as they were not considered reliable (sample sizes was very small for certain cohorts, and in several instances poor covariate balancing was achieved following the implementation of the PSM approach).

Table C.7 Sample sizes available for the impact analysis based on administrative data, by project

Pilot project	2012/13 cohort	2013/14 cohort	2014/15 cohort
APM Ltd	125	∅	∅
Energise Ltd	320	367	246
Links to Success	∅	81	∅
PEF	∅	∅	∅
Prevista	126	∅	∅
Teens & Toddlers	267	214	134
Triodos	345	192	148
<i>Base: 2,817 respondents.</i>			

⁵The variables reflecting an individual’s previous Key Stage (KS) 2 attainment in English are three binary indicators denoting whether the pupil achieved a national curriculum level below, at or above the expected level for KS2, which is a national curriculum level 4. The variables indicating an individual’s KS2 attainment in Mathematics were obtained in a similar way.

⁶ The Index of Multiple Deprivation (IMD) score of the area (Lower Super Output Area) in which the last school attended by the pupil was located. Higher scores indicate greater levels of deprivation.

C.3 Representativeness of the samples and sample size considerations

C.3.1 Impact analysis based on survey data

One possible disadvantage of using survey data to assess impacts stems from the fact that survey respondents in the participant group (908 individuals) make up only 29.7 per cent of all participants observed in the MI (3,055 individuals at the time of the survey-based impact analysis), and therefore they may not be representative of the entire population of those who participated in the IF pilot.

Unfortunately, it was not possible to explore whether any differences existed between the 2,189 individuals issued for the survey, who could be linked to the NPD, and the 866 unlinked participants as the characteristics of the latter could not be observed. Therefore, we simply have to assume that the initial stage performed to select the issued sample produced a representative sample of participants. Nevertheless, it is worth noting that the 2,189 individuals issued for the survey made up 71.7 per cent of the participant population.

The other potential source of bias is related to survey response (or non-response) within the issued sample: if those who responded to the survey were different from those who did not, relying on respondents to make inference on the population of all participants would be misleading. Comparing some of the basic characteristics (e.g. age, gender and FSM eligibility and level of persistent absenteeism) of IF participants who responded to the survey (908 cases) against the same characteristics for those IF participants who did not respond (1,281 cases were available), it was possible to construct non-response weights. However, when we tested whether using weighted instead of unweighted data made any difference for the analysis, we found that this was not the case. The composition of the IF participant sample observed using unweighted data was in fact the same as the composition observed using data adjusted with non-response weights: virtually all the IF participant sample's features listed in Tables C.1 to C.3 (demographic, educational and parents/family related characteristics) did not vary following the use of non-response weights. For this reason, the analysis used unweighted figures.

A final consideration pertains to the ability to detect statistically significant results using survey data, as this may be limited by the available sample sizes. Fewer cases for analysis generally means that the impact of the IF, if it exists, is less likely to be detected (technically, to be found to be statistically significant). However, this disadvantage is to some extent offset by the rich nature of the survey data as this data allow for more individual characteristics to be taken into account, thereby increasing the precision of impact estimates.

⁷ The only exception was the proportion of IF participants who were 17 years old at the time they started the IF: this proportion was 18.7 per cent using weighted data and 14.7 per cent using non-response weights. The difference between these two proportions was found to be statistically significantly different from zero at the 5 per cent level of significance. This means that the proportion of 17 years old in the population of IF participants was four percentage points higher than in the survey sample. However, note that non-response weights were constructed using a limited number of variables (mainly variables available from the NPD prior to conducting the survey of participants) and might have therefore provided limited information to ensure that IF participants who answered the survey were representative of the whole population of programme participants.

C.3.2 Impact analysis based on administrative data

The impact analysis which used administrative data was based on a sample of 2,817 IF participants, all of whom were aged 14 or 15 at the programme start and began the programme within any of the time periods among November 2012-April 2013, November 2013-April 2014 or November 2014-April 2015. While these individuals made up over half (52 per cent) of the population of IF participants of the same age range who could be matched to NPD/ILR data (5,597 young people in total), it is possible that they do not reflect the general profile of this population. Understanding whether our analytical sample is representative of the IF participant population is crucial to generalise the findings of this report to all disadvantaged people covered by the IF provision.

Table C.8 compares our analytical sample of 2,817 participants and the remainder of participants not included in the analysis (2,700 individuals, including September-October and May-August starters) across all the baseline characteristics considered. The figures suggest that participants in the unused sample were likely to be more disadvantaged than those in the analytical sample (e.g., they were more likely to be eligible for FSM and have a KS2 Maths attainment below the expected level, a larger proportion were excluded from school and they also showed higher values of the Index of Multiple Deprivation). The implications of leaving this group of participants out of the analysis are unclear as they might have either benefited from the programme or proved to be harder to support, with an ambiguous effect on (i.e., under- or over-estimation of) the impact estimates.

Table C.8 Comparison of the analytical sample (all cohorts) with the unused sample (remainder of participants) aged 14-15.

	Analytical sample (all cohorts) %	Unused sample (rest of participants) %
Aged at IF start		
14 (%)	57.2***	63.1
15 (%)	42.8***	36.9
Gender		
Female (%)	50.2***	45.1
Male (%)	49.8***	54.9
Ethnicity		
White (%)	77.7***	66.4
Black (%)	7.4***	10.4
Asian (%)	7.6***	15.1
Mixed (%)	5.6	6.1
Other (%)	1.7	2.0
English as a first language (%)	93.6***	86.2
Special Education Needs (%)	33.2	31.0
Eligible for Free School Meals (%)	43.2***	48.8
KS2 English attainment		
Below expected level (%)	40.1	41.5
At expected level (%)	49.5	49.1
Above expected level (%)	10.4	9.4
KS2 Maths attainment		
Below expected level (%)	30.8***	33.4
At expected level (%)	58.4	56.9
Above expected level (%)	10.8	9.8
Persistently absent (%)	5.4	4.7
Excluded (%)	15.7***	17.8

D Experiences of participants and non-participants who responded to the survey

D.1 Experiences of IF pilots and other programmes and support

Exploring survey data provided information about young people's experiences of taking part in the IF (Round One) projects, and of participating in other local area programmes and/or receiving other forms of supports beyond the pilot. Individuals who participated in the IF projects were asked a number of questions about this subject, including about the type of support they received from the pilot, how they heard about it, the frequency and length of the sessions, and the reason that they participated in the intervention. Experiences of participation in programmes other than the IF pilot and of receiving other forms of support were also explored for those who did not take part in the IF.

D.1.1 Experiences of the IF pilot

Support received

The Round One projects provided support to young people in a number of different ways. These included, for example, support from a personal advisor acting in a coaching or mentoring role, help to develop job search and application skills, and provision of opportunities to study for qualifications, access training or obtain work experience placements. IF participants were asked whether they had received any type of support from their specific IF project (the exact name of the project was provided by the interviewers). They could choose among different types of support listed in a show-card which was handed to them.

Interestingly, only two fifths of the participant sample (41.3 per cent of the 891 respondents, which amounts to 368 young people) reported having received any form of support. The other three fifths (523 individuals, almost 60 per cent of the participant sample) did not mention having received any form of support from the project in which they participated. Table D.1 shows the proportions, among IF participants who claimed to have received at least one form of support (368 individuals), that mentioned each form of support. Note that each individual could mention more than one form of support and therefore the types reported in the table are not mutually exclusive (this is the reason that the sum of the percentages exceeds 100).

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

Table D.1 Types of support received by IF participants

Type of support	%
Personal advisor	47.0
Job application skills	43.8
Work experience	35.9
Career guidance	34.2
Job search skills	32.6
Accessing training, education or apprenticeships	31.3
Basic skills course	23.9
Practical help	21.7
Organised group activity	15.5
Other organised activities	15.5
Other activities	4.6
Emotional support or counselling	1.4
Help with literacy and/or numeracy skills	1.1
Confidence building	0.5

Base: 368 respondents

As the table shows, almost half (47 per cent) of those IF participants who claimed to have received support reported that they have been helped by a personal advisor, and this was the most frequently cited form of support. This result is in line with the findings from the qualitative evaluation of early implementation of the IF, according to which participants’ relationship with their personal advisor was a key defining feature of the IF pilot projects and was particularly valued by young people.

Other frequently mentioned forms of support were those relating to the development of work-related skills: job application skills was mentioned by 43.8 per cent of the 368 IF participants, and around one third of these 368 individuals mentioned having received support on work experience, career guidance and job search skill (35.9, 34.2 and 32.6 per cent, respectively). Notably, almost one third (31.3 per cent) of those who claimed to have received any form of support mentioned this related to accessing training, education or apprenticeships. Other forms of support were also reported, even if less frequently than those mentioned above. For example, a quarter of the 368 IF participants cited having been provided with a basic skills course and one fifth were offered practical help.

The finding that only 368 of the 891 IF participants who responded to the survey mentioned some form of support has several possible explanations. For example, it is possible that the majority of participants did not engage with the projects, and as a consequence their reports of having received no support are genuine ones. However, this justification per se does not seem convincing as the qualitative study by Griffiths et al. (2016) emphasised the high level of support that young people generally reported that they had received. An alternative explanation could be that providers might have deliberately downplayed the support they were giving to reduce the risk of stigmatising young people.

It can also be speculated that poor recollection of the projects played a role in the findings of under-reporting of the support received. For example, it is possible that participants could not link the support received to a particular project as some of the projects were known under several names and this could have been misleading. For example, Triodos is also known as New Horizons or Career Connect. Furthermore, the building in which the project is delivered has the Connexions logo in several places (Connexions is a service that provides career advice and guidance for young people). While interviewers specifically mentioned the name of the IF project in which the respondent participated when asking the question about support (thus reducing the likelihood of recollection bias), it is still possible that the different names by which some of the projects were known caused some confusion.

Recollection could have been scarcer among younger IF participants (those aged 14 or 15) than older ones. The qualitative evaluation of the IF pilot found that most projects were taking place at school, during the school day and with school-aged people. For these participants, support could have easily blurred into 'school as usual.' On the other hand, for people who had left school participation was expected to be much more active. There was no teacher to refer them and therefore they had to make a conscious decision to refer themselves. These people either had to travel to the delivery body or have the delivery body visit their home, and therefore their awareness of the support received by the IF pilot is expected to be greater than for younger participants. However, data checks indicated that the oldest participants (those aged 16 or over at interview) reported having received support less frequently than participants aged 14 or 15, undermining the plausibility of this explanation.

It is possible that most of IF participants were not helped to the extent that they felt they needed, and this prompted them to fail to report that they received support in any form from the IF projects. Another explanation is that the support that young people received from IF pilot has been seen as another aspect of the alternative support that they have accessed from other sources. Table D.2 shows that around two-fifths (40.1 per cent) of the IF participant sample acknowledged the receipt of some form of non-IF related support.

D.1.2 Experiences of other programmes or support

Experiences of both IF participants and non-participants with regard to other programmes and support not related to the IF pilot were explored. In order to evaluate the success of the IF pilot projects it is indeed crucial to understand the extent to which these projects added value to the provision which would have otherwise been available to young people.

In areas where the IF pilot was delivered there were other national programmes operating. For example, the Youth Contract, the Youth Employment Initiative, the Youth Engagement Fund and the European Social Fund targeted young people similar to those supported by the IF. In addition, a number of schemes provided by voluntary organisations, local authorities and individual schools operated at the local area level. The qualitative evaluation of the IF pilot noted that programme providers had to actively market their services to schools against multiple competing offers of support for young people.

Establishing the nature and extent of support received by young people in addition or alternatively to the IF pilot can provide crucial information to interpret young people's outcomes observed after participation in the IF, and therefore help contextualise the estimated impacts reported in the next section.

Support received

In the survey, young people were asked about the help or support that they received in the last twelve months from any programme or service, or from one-to-one support from a mentor (excluding personal help or support from a relative or friend and one-off help from teachers). Survey respondents were asked to choose the show-cards which best described this support, and IF participants were explicitly told that the support in question did not refer to the IF pilot.

Non-IF support was more frequently mentioned among the 891 matched non-participants compared to the 891 IF participants. Around two-fifths (40.1 per cent, which amounts to 357 individuals) of the IF participant sample reported having

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

received some form of non-IF support. A much higher proportion, around two-thirds (67.2 per cent, 599 individuals) of the matched non-participant sample mentioned to have received non-IF support.

The proportion of young people who received some type of non-IF related support is reported in Table D.2, separately for IF participants and the matched non-participant sample. (Note that respondents could mention more than one type of support, and therefore figures sum to more than 100.)

Table D.2 Types of non-IF support received in the last 12 months

Type of support	IF participants	Matched non-participants
	(%)	(%)
Personal advisor	17.0	14.4
Job application skills	14.3	19.5
Work experience	10.7	22.5
Career guidance	11.5	17.6
Job search skills	10.1	13.1
Access to training, education or apprenticeships	9.8	8.3
Basic skills course	6.9	8.0
Practical help	6.7	8.2
Organised group activity	3.8	7.2
Other organised activities	3.2	7.9
Any other support or activities	1.7	1.0

Base: 956 respondents (357 IF participants and 599 matched non-participants)

As the table shows, with the exception of three support types (namely, Personal advisor, Access to training, education and apprenticeships, and the residual category Any other support or activities), all other non-IF support types were mentioned by a higher proportion of young people in the matched non-participant sample compared to the IF participant sample. For example, the proportion of young people who mentioned having received support on work experience in the matched comparison group was almost twice as large as in the IF participant group (22.5 and 10.7 per cent, respectively). Support to develop job application skills was mentioned by one in five young people (19.5 per cent) in the matched non-participant sample but only one in seven (14.3 per cent) IF participants.

Sources of information

Table D.3 shows how IF participants and non-participants who received non-IF related support heard about it.

Table D.3 Source of knowledge about the non-IF programme or support

Source of knowledge	IF participants	Matched non-participants
Job centre	6.4	4.3
School or college	62.5	72.0
Youth offending team or probation court	2.8	0.5
Another local community organisation	2.0	2.8
Social worker	3.1	1.0
Someone else in the community	2.0	1.0
Friend or family member	10.6	11.2
Advert	1.1	0.8
Internet or online	2.2	1.8
Spoken to or approached by the organisation	0.6	0.3
Connexions	1.4	0.3
Other	4.5	3.0
Not answered	0.9	0.8

Base: 956 respondents (357 IF participants and 599 matched non-participants)

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

Among the survey respondents who acknowledged having received non-IF support (357 IF participants and 599 non-participants), a school or college was the most frequently mentioned source of information about the support: 62.5 per cent of the IF participants and 72 per cent of non-participants mentioned this source. Around 11 per cent of IF participants mentioned having heard about other support or another programme from a friend or family member, and a similar proportion among non-participants mentioned this source. Around one in every twenty individuals (6.4 and 4.3 per cent of those in the IF and non-participant sample, respectively) mentioned that the source of knowledge of the programme/support was a job centre.

Frequency of the support

Table D.4 shows how frequently IF participants and non-participants reported to have received non-IF support. According to the survey reports, support once a month or more often (indicated by the first four answers listed in the table) was cited more often among IF participants than their matched counterparts. Over a third (34.4 per cent) of non-participants reported that it was one-off support, which compares to one-sixth (16.5 per cent) for IF participants.

Table D.4 Frequency of non-IF support

Frequency	IF participants	Matched non-participants
	%	%
Every day	8.1	3.2
Two to three times a week	13.7	11.7
Once a week	27.2	21.0
Less often than once a week but at least once a month	21.0	13.9
Less than once a month	12.6	15.5
One-off support	16.5	34.4
Not answered	0.8	0.3

Base: 956 respondents (357 IF participants and 599 matched non-participants)

Length of the sessions and attendance

Figures reported in Table D.5 show a similar proportion of IF participants and non-participants (nearly 40 per cent) reported that the sessions of the non-IF programme/support lasted less than an hour. A larger proportion experienced sessions of length between one and four hours among IF participants, while non-participants were more likely to state that they attended sessions longer than four hours.

Table D.5 Length of non-IF sessions

Session length	IF participants	Matched non-participants
	%	%
Less than an hour	37.3	38.7
One to two hours	37.0	29.9
Three to four hours	7.6	4.7
More than four hours	13.5	20.4
Length of session varies	4.2	5.7
Not answered	0.6	0.7

Base: 956 respondents (357 IF participants and 599 matched non-participants)

As Table D.6 indicates, a higher proportion reported having missed some sessions among IF participants (21.9 per cent) compared to matched non-participants (15.2 per cent).

Table D.6 Whether individuals ever missed any session

Survey response	IF participants	Matched non-participants
	%	%
Yes	21.9	15.2
No	77.9	84.6
Not answered	0.3	0.2

Base: 956 respondents (357 IF participants and 599 matched non-participants)

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

Decision to participate and willingness to continue

According to the reports (see Table D.7), almost three quarters (73.3 per cent) of non-participants who received non-IF related support stated that they were willing to receive such support. A lower proportion (63.9 per cent) was observed among IF participants. The latter were also more likely to report that receiving support was on someone else's initiative.

Table D.7 Whether individuals decided themselves to take part in the non-IF programme/support

Survey response	IF participants	Matched non-participants
	%	%
Wanted to receive support	63.9	73.3
Made to receive support	33.9	25.7
Other	1.9	0.7
Not answered	0.3	0.3

Base: 956 respondents (357 IF participants and 599 matched non-participants)

Around three-fifths (60.8 per cent) of IF participants who reported having received support said that they would like to continue receiving it, and a similar proportion (59.6 per cent) was observed among matched non-participants (see Table D.8).

Table D.8 Whether individuals would like to continue with the non-IF programme/support

Survey response	IF participants	Matched non-participants
	%	%
Yes	60.8	59.6
No	37.8	38.1
Not answered	1.4	2.3

Base: 956 respondents (357 IF participants and 599 matched non-participants)

Perceived impact

As Table D.9 shows, the vast majority of young people who stated that they received non-IF related support reported that the support was helpful (their response being either Very helpful or Fairly helpful), and this finding held for both IF participants and those in the matched group of non-participants.

Table D.9 How helpful the non-IF programme/support was overall

Survey response	IF participants	Matched non-participants
	%	%
Very helpful	52.1	45.1
Fairly helpful	34.7	42.2
Made no difference	8.4	8.2
Fairly unhelpful	2.0	2.5
Very unhelpful	2.5	2.0
No response	0.3	0.0

Base: 956 respondents (357 IF participants and 599 matched non-participants)

Among those who said that the non-IF support was helpful (310 IF participants and 532 matched non-participants), high proportions reported that the support helped them better understand what work is like, improve self-confidence, motivation and communication with other people (see Table D.10). The benefits cited were between 68.8 and 85.5 per cent. IF participants were more likely than matched non-participants to report each of these benefits.

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

Very few young people answered the question about whether the non-IF support helped them improve the frequency of school or college attendance. Non-reporting was more commonly observed among IF participants (88 per cent did not provide an answer) than matched non-participants (70 per cent). IF participants were more likely than non-participants to mention that non-IF support helped them gain qualifications (44.8 and 41.4 per cent).

Table D.10 How the IF was helpful to participants

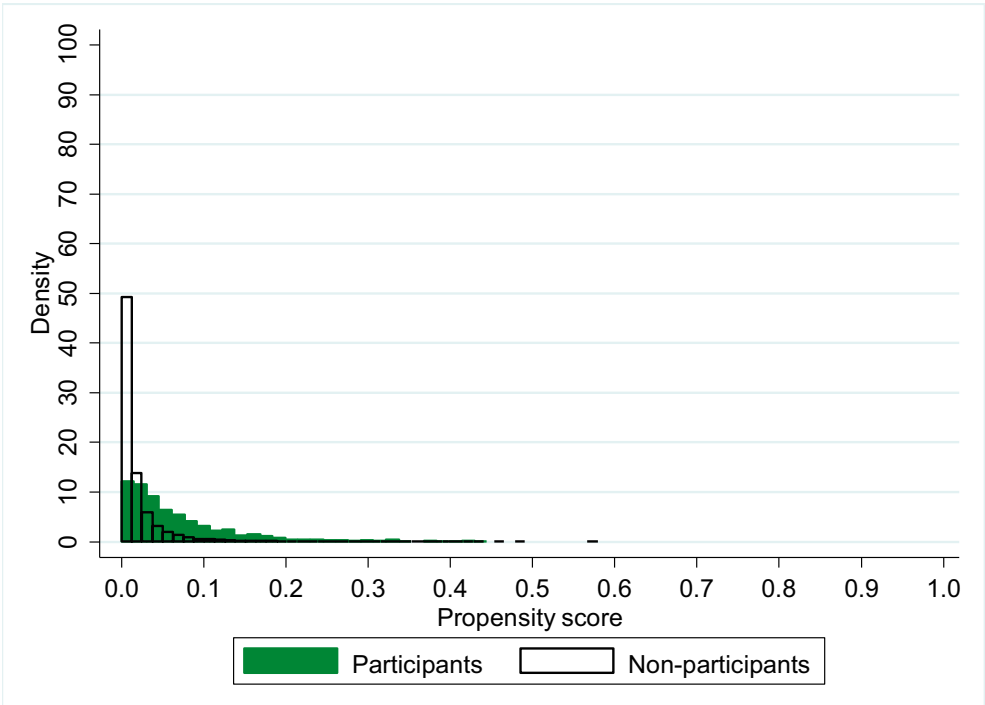
Survey response	Yes	No	Not answered
	%	%	%
Get a better understanding of what working is like			
IF participants	73.2	17.4	9.4
Matched non-participants	68.8	18.8	11.8
Improve self-confidence			
IF participants	79.0	19.4	1.6
Matched non-participants	78.0	19.7	0.6
Improve motivation			
IF participants	85.5	13.9	0.6
Matched non-participants	82.5	15.2	0.6
Improve communication with other people			
IF participants	80.3	19.0	0.6
Matched non-participants	74.4	23.9	0.0
Improve frequency of attendance of school or college			
IF participants	4.8	7.1	88.1
Matched non-participants	7.9	19.7	70.7
Gain qualifications			
IF participants	44.8	54.2	1.0
Matched non-participants	41.4	56.2	0.8
<i>Base: 842 respondents (310 participants and 532 non-participants)</i>			

E Distribution of the propensity score

The graphs displayed in this section of the appendix show the frequency distribution of the propensity score (i.e., the estimated probability of participating in the IF pilot), prior to implementing the propensity score matching approach, for the samples of IF participants and non-participants. Each frequency distribution is represented by means of rectangles. The width of each rectangle represents the class interval and its height the average frequency density for the interval.

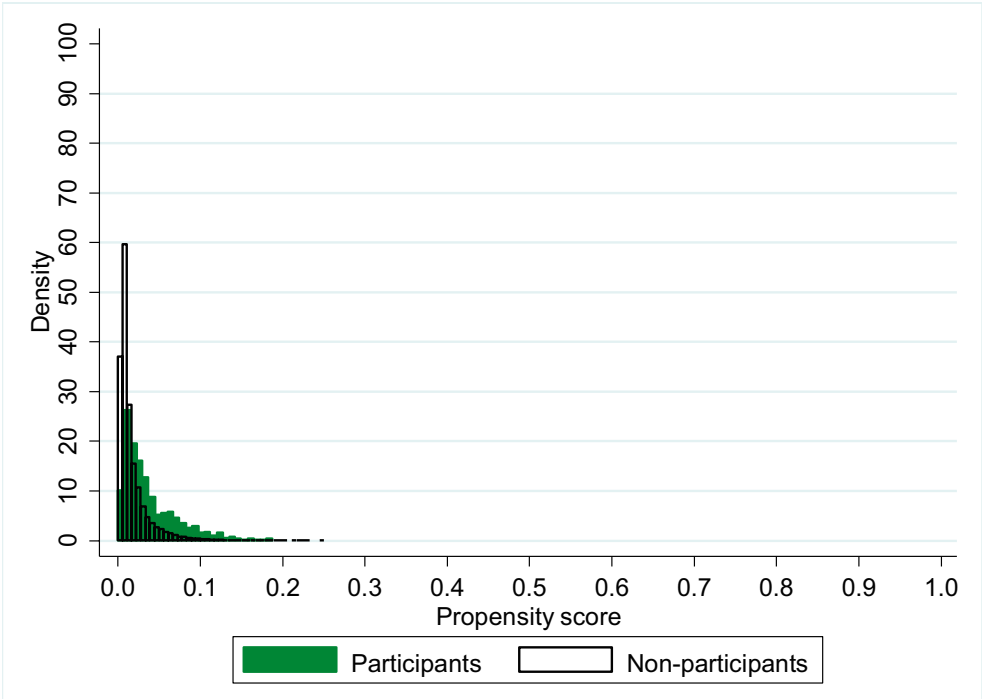
All of the distributions are skewed towards zero and, in general, there are very few scores with positive values above 0.2, which is suggestive of poor predictors of participation in the IF. However, the large size of the non-participant group relative to the participant group, makes it difficult to establish whether this is the case, as the estimated scores are artificially closer to zero than they would be in the presence of a smaller group on non-participants.

Figure E.1 Distribution of the propensity score for IF participants and non-participants (2012/13 cohort, educational outcomes)



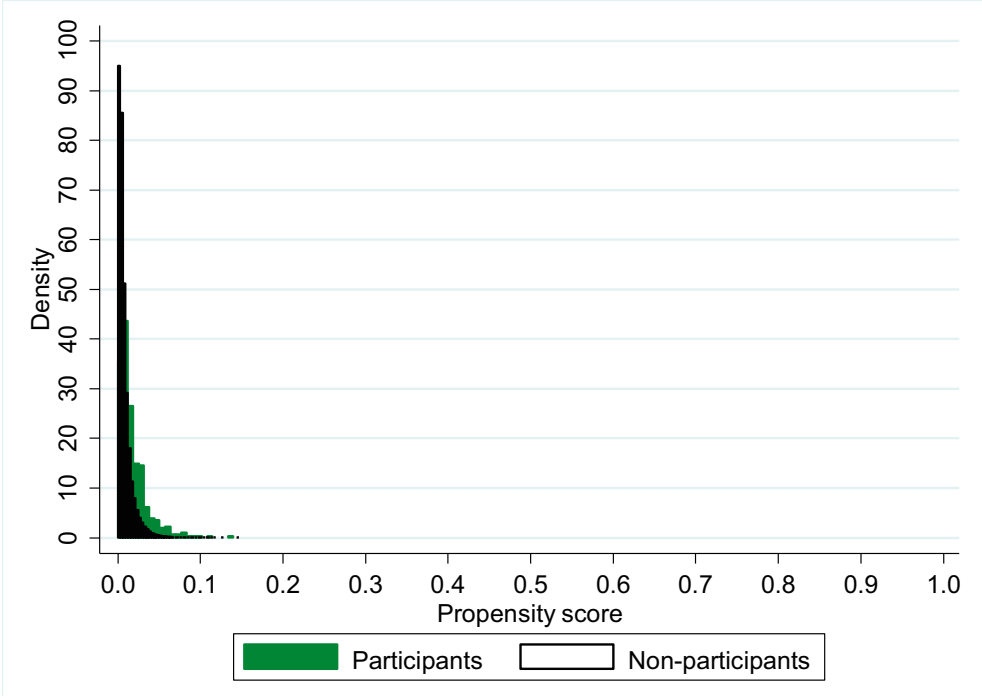
Base: 889 IF participants and 39,835 non-participants

Figure E.2 Distribution of the propensity score for IF participants and non-participants (2013/14 cohort, educational outcomes)



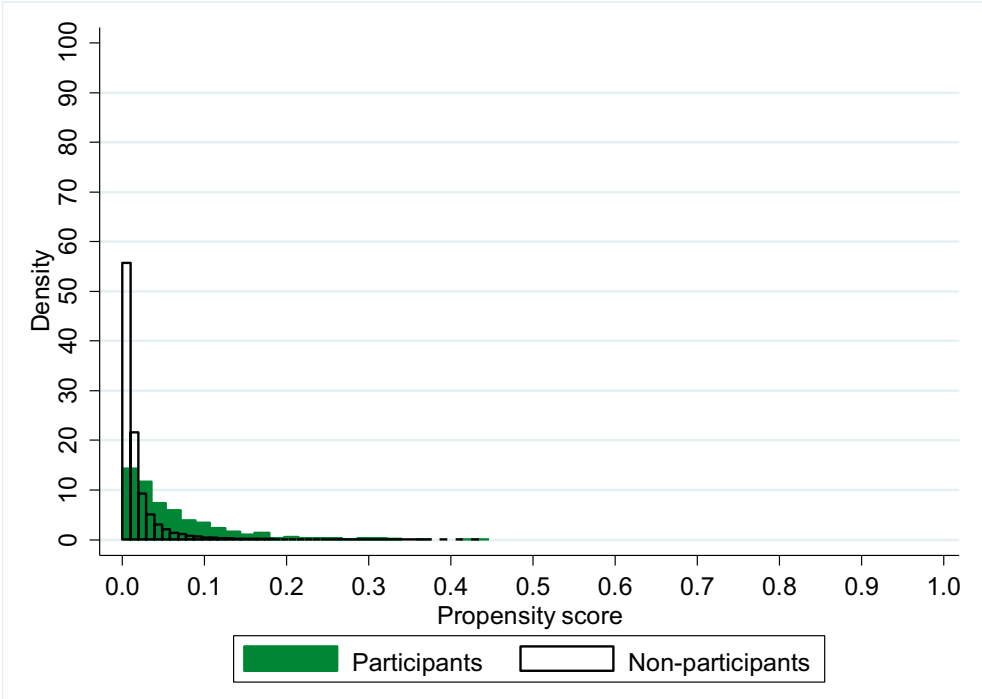
Base: 621 participants and 32,806 non-participants

Figure E.3 Distribution of the propensity score for IF participants and non-participants (2014/15 cohort, educational outcomes)



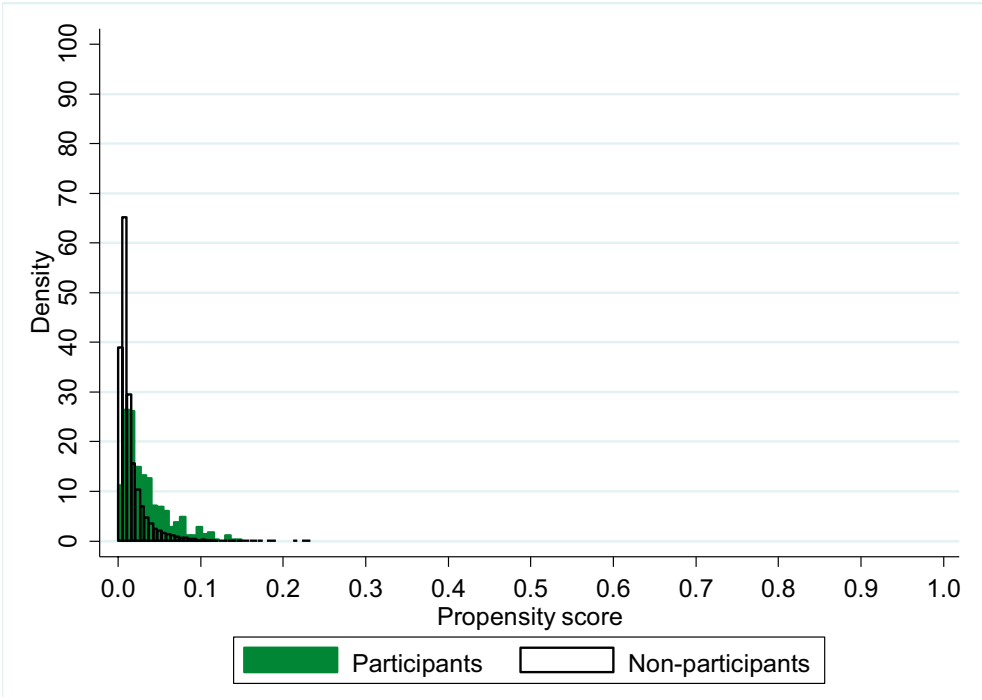
Base: 462 participants and 50,153 non-participants

Figure E.4 Distribution of the propensity score for IF participants and non-participants (2012/13 cohort, persistent absenteeism)



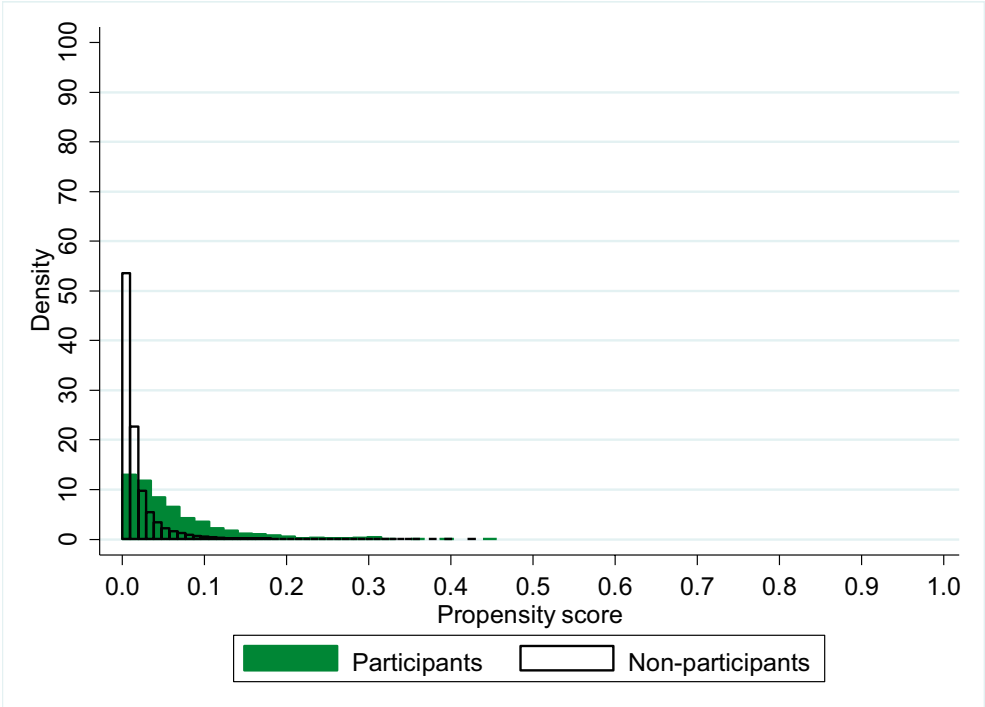
Base: 613 participants and 29,978 non-participants

Figure E.5 Distribution of the propensity score for IF participants and non-participants (2013/14 cohort, persistent absenteeism)



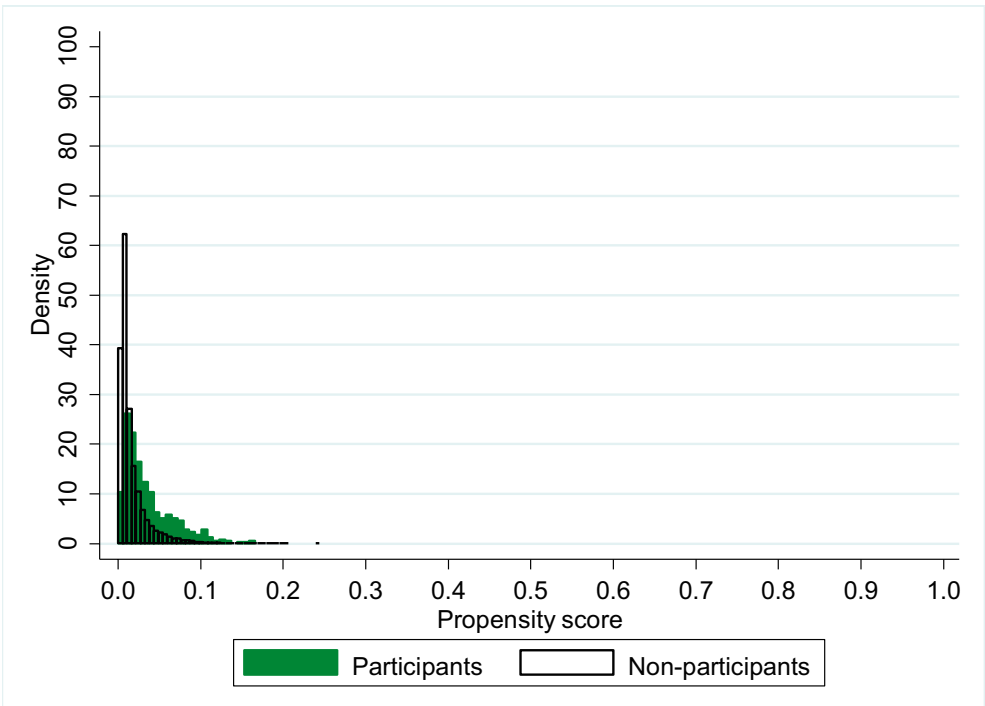
Base: 481 participants and 27,653 non-participants

Figure E.6 Distribution of the propensity score for IF participants and non-participants (2012/13 cohort, exclusions)



Base: 651 participants and 30,584 non-participants

Figure E.7 Distribution of the propensity score for IF participants and non-participants (2013/14 cohort, exclusions)



Base: 513 participants and 28,384 non-participants

F By-round impact analysis (additional statistics)

Table F.1 Proportions of the participant and comparison group who achieved a first NQF level 1 qualification at each time point between the end of the IF start year and three years after, by round of commissioning

	2012/13 cohort		2013/14 cohort		2014/15 cohort	
	Participants	Comparison	Participants	Comparison	Participants	Comparison
	%	%	%	%	%	%
Round One						
End of academic year (0)	28.6	29.1	20.6	17.3	35.9	34.8
One year after (1)	89.0	82.3	77.0	75.5	89.8	68.1
Two years after (2)	95.2	94.3	83.1	87.4		
Three years after (3)	95.2	94.5				
Round Two						
End of academic year (0)	18.9	19.6	12.5	12.9	26.3	26.4
One year after (1)	77.5	76.9	83.5	79.4	65.8	58.4
Two years after (2)	93.6	93.0	96.7	93.9		
Three years after (3)	95.1	93.8				
<i>Base (Round One): 419 participants and 398 comparators (2012/13 cohort), 284 participants and 268 comparators (2013/14 cohort), and 128 participants and 126 comparators (2014/15 cohort).</i>						
<i>Base (Round Two): 530 participants and 513 comparators (2012/13 cohort), 425 participants and 427 comparators (2013/14 cohort), and 357 participants and 373 comparators (2014/15 cohort).</i>						

Table F.2 Proportions of the participant and comparison group who achieved a first NQF level 2 qualification at each time point between the end of the IF start year and three years after, by round of commissioning

	2012/13 cohort		2013/14 cohort		2014/15 cohort	
	Participants	Comparison	Participants	Comparison	Participants	Comparison
	%	%	%	%	%	%
Round One						
End of academic year (0)	2.9	6.3	2.1	4.9	3.9	4.4
One year after (1)	18.4	26.0	15.2	26.5	26.6	23.7
Two years after (2)	33.7	44.7	21.8	33.3		
Three years after (3)	34.4	45.7				
Round Two						
End of academic year (0)	2.5	6.0	2.8	3.9	0.0	1.5
One year after (1)	12.1	30.2	15.5	30.2	7.3	15.9
Two years after (2)	31.1	47.9	20.7	43.1		
Three years after (3)	31.5	49.1				
<i>Base (Round One): 419 participants and 398 comparators (2012/13 cohort), 284 participants and 268 comparators (2013/14 cohort), and 128 participants and 126 comparators (2014/15 cohort).</i>						
<i>Base (Round Two): 530 participants and 513 comparators (2012/13 cohort), 425 participants and 427 comparators (2013/14 cohort), and 357 participants and 373 comparators (2014/15 cohort).</i>						

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

Table F.3 Proportions of the participant and comparison group who achieved their first NQF level 3 qualification at each time point between the end of the IF start year and three years after, by round of commissioning

	2012/13 cohort		2013/14 cohort		2014/15 cohort	
	Participants	Comparison	Participants	Comparison	Participants	Comparison
	%	%	%	%	%	%
Round One						
End of academic year (0)	0.0	0.0	0.0	0.0	0.0	0.0
One year after (1)	1.4	3.1	0.0	1.2	0.0	0.0
Two years after (2)	4.8	14.2	0.8	4.5		
Three years after (3)	6.7	19.2				
Round Two						
End of academic year (0)	0.0	0.0	0.0	0.0	0.0	0.0
One year after (1)	0.4	1.5	0.0	0.3	0.0	0.0
Two years after (2)	6.6	10.2	0.5	2.1		
Three years after (3)	8.3	20.6				

Base (Round One): 419 participants and 398 comparators (2012/13 cohort), 284 participants and 268 comparators (2013/14 cohort), and 128 participants and 126 comparators (2014/15 cohort).

Base (Round Two): 530 participants and 513 comparators (2012/13 cohort), 425 participants and 427 comparators (2013/14 cohort), and 357 participants and 373 comparators (2014/15 cohort).

Table F.4 Proportions of the participant and comparison group who were persistently absent and excluded from school in the academic year after starting the pilot, by round of commissioning

	2012/13 cohort	2013/14 cohort
Round One		
Persistently absent		
Participants (%)	11.2	17.3
Comparison (%)	10.0	9.3
Excluded		
Participants (%)	13.9	15.7
Comparison (%)	9.8	12.9
Round Two		
Persistently absent		
Participants (%)	12.3	7.0
Comparison (%)	7.7	(4.2)
Excluded		
Participants (%)	18.1	12.5
Comparison (%)	9.9	6.5

Base (Round One): 250 participants and 221 comparators (Persistently absent), and 281 participants and 260 comparators (Excluded) for the 2012/13 cohort; 150 participants and 150 comparators (Persistently absent), and 178 participants and 176 comparators (Excluded) for the 2013/14 cohort.

Base (Round Two): 414 participants and 395 comparators (Persistently absent), and 425 participants and 410 comparators (Excluded) for the 2012/13 cohort; 362 participants and 356 comparators (Persistently absent), and 371 participants and 375 comparators (Excluded) for the 2013/14 cohort.

**** and **: Statistically significant at the 1 and 5 per cent level, respectively.*

Table F.5 Difference between Round One and Round Two impacts on the likelihood of being persistently absent and excluded from school in the academic year after starting the pilot

	2012/13 cohort	2013/14 cohort
Persistently absent (impact in ppts)	-3.4	5.2
Excluded (impact in ppts)	-4.1	-3.2

G By-gender impact analysis (additional statistics)

Table G.1 Proportions of the participant and comparison group who achieved their first NQF level 1 qualification at each time point between the end of the IF start year and three years after, by gender of young person

	2012/13 cohort		2013/14 cohort		2014/15 cohort	
	Participants	Comparison	Participants	Comparison	Participants	Comparison
	%	%	%	%	%	%
Male						
End of academic year (0)	24.0	25.3	20.1	19.4	31.2	32.4
One year after (1)	82.6	78.5	79.7	75.1	79.0	65.4
Two years after (2)	93.6	90.7	89.1	90.3		
Three years after (3)	94.4	90.7				
Female						
End of academic year (0)	22.4	15.7	10.0	11.8	27.1	27.9
One year after (1)	82.6	77.3	82.8	73.1	67.1	65.8
Two years after (2)	95.1	93.5	94.8	88.8		
Three years after (3)	95.9	94.1				
<i>Base (Male): 484 participants and 436 comparators (2012/13 cohort), 359 participants and 339 comparators (2013/14 cohort), and 205 participants and 204 comparators (2014/15 cohort).</i>						
<i>Base (Female): 465 participants and 455 comparators (2012/13 cohort), 309 participants and 301 comparators (2013/14 cohort), and 280 participants and 283 comparators (2014/15 cohort).</i>						

Table G.2 Proportions of the participant and comparison group who achieved their first NQF level 2 qualification at each time point between the end of the IF start year and three years after, by gender of young person

	2012/13 cohort		2013/14 cohort		2014/15 cohort	
	Participants	Comparison	Participants	Comparison	Participants	Comparison
	%	%	%	%	%	%
Male						
End of academic year (0)	2.3	5.3	3.1	3.3	0.0	2.9
One year after (1)	12.0	21.4	12.3	22.0	9.8	25.4
Two years after (2)	27.1	37.7	14.8	29.1		
Three years after (3)	27.7	38.4				
Female						
End of academic year (0)	3.0	5.6	1.9	4.2	1.8	1.8
One year after (1)	18.1	29.4	19.1	28.5	14.3	26.1
Two years after (2)	37.6	49.3	28.5	40.5		
Three years after (3)	38.1	51.4				
<i>Base (Male): 484 participants and 436 comparators (2012/13 cohort), 359 participants and 339 comparators (2013/14 cohort), and 205 participants and 204 comparators (2014/15 cohort).</i>						
<i>Base (Female): 465 participants and 455 comparators (2012/13 cohort), 309 participants and 301 comparators (2013/14 cohort), and 280 participants and 283 comparators (2014/15 cohort).</i>						

Evaluation of the Innovation Fund pilot: Quantitative assessment of impact and social return on investment

Table G.3 Proportions of the participant and comparison group who achieved their first NQF level 3 qualification at each time point between the end of the IF start year and three years after, by gender of young person

	2012/13 cohort		2013/14 cohort		2014/15 cohort	
	Participants	Comparison	Participants	Comparison	Participants	Comparison
	%	%	%	%	%	%
Male						
End of academic year (0)	0.0	0.0	0.0	0.0	0.0	0.0
One year after (1)	0.8	1.7	0.0	0.8	0.0	0.0
Two years after (2)	3.7	9.5	1.1	1.9		
Three years after (3)	4.8	14.2				
Female						
End of academic year (0)	0.0	0.0	0.0	0.0	0.0	0.0
One year after (1)	0.9	1.5	0.0	0.2	0.0	0.0
Two years after (2)	8.0	10.2	0.0	1.5		
Three years after (3)	10.5	21.9				
<i>Base (Male): 484 participants and 436 comparators (2012/13 cohort), 359 participants and 339 comparators (2013/14 cohort), and 205 participants and 204 comparators (2014/15 cohort).</i>						
<i>Base (Female): 465 participants and 455 comparators (2012/13 cohort), 309 participants and 301 comparators (2013/14 cohort), and 280 participants and 283 comparators (2014/15 cohort).</i>						

Table G.4 Difference between impacts for male and female participants on the likelihood of achieving a first NQF level 3 qualification

	End of year	One year after	Two years after	Three years after
2012/13 cohort (impact in ppts)	0.0	-0.2	-3.6	1.9
2013/14 cohort (impact in ppts)	0.0	-0.7	0.7	
2014/15 cohort (impact in ppts)	0.0	0.0		

Table G.5 Proportions of the participant and comparison group who were persistently absent and excluded from school in the academic year after starting the pilot, by gender of participant

	2012/13 cohort	2013/14 cohort
Male		
Persistently absent		
Participants (%)	14.6	8.8
Comparison (%)	12.1	4.2
Excluded		
Participants (%)	20.5	15.1
Comparison (%)	15.2	14.9
Female		
Persistently absent		
Participants (%)	14.3	11.0
Comparison (%)	14.3	5.5
Excluded		
Participants (%)	7.3	12.0
Comparison (%)	7.3	6.7
<i>Base (Male): 321 participants and 287 comparators (Persistently absent), and 352 participants and 316 comparators (Excluded) for the 2012/13 cohort; 238 participants and 231 comparators (Persistently absent), and 271 participants and 255 comparators (Excluded) for the 2013/14 cohort.</i>		
<i>Base (Female): 343 participants and 325 comparators (Persistently absent), and 354 participants and 335 comparators (Excluded) for the 2012/13 cohort; 272 participants and 261 comparators (Persistently absent), and 276 participants and 267 comparators (Excluded) for the 2013/14 cohort.</i>		