Department for Education

Taking Teaching Further Financial Incentive: impact evaluation feasibility study

Research report

September 2024

Author: Verian

Government Social Research

This research report was written before the new UK government took office on 5 July 2024. As a result, the content may not reflect current government policy.

Contents

List of abbreviations

AoC: Association of Colleges

DfE: Department for Education

DiD: Difference in difference

ETF: Education and Training Foundation

FE: Further education

FEWDC: Further education workforce data collection

FI: Financial incentive

FTE: Full-time equivalent

FY: Financial year

ITP: Independent training provider

MDES: Minimum detectable effect size

RCT: Randomised controlled trial

RDD: Regression discontinuity design

RQ: Research question

SDiD: Synthetic difference in difference

SEND: Special education needs and disabilities

ToC: Theory of Change

TTF: Taking Teaching Further

UK: United Kingdom

1. Executive summary

Aims and Objectives

Taking Teaching Further (TTF) supports further education (FE) providers to recruit those with relevant knowledge and industry experience to retrain as FE teachers. Previous rounds have given funding to support providers, constituting the 'core' TTF offer. In the sixth round (launched in April 2023), in addition to the 'core' TTF offer, a pilot of a new financial incentive (FI) was introduced for TTF recruits in some of the most hard-to-fill subject areas. The FI gives recruits £3,000 at the end of each year of the TTF programme. The 'FI subjects' are: digital; construction and the built environment; engineering and manufacturing; and maths.

The Department for Education (DfE) commissioned Verian to assess the feasibility of an impact evaluation of the FI, which would answer the following three key research questions:

1) Does the introduction of the FI significantly increase teacher recruitment in FE settings?
2) Does the introduction of the FI significantly increase teacher retention in FE settings?
3) What, if any, are the unintended outcomes and negative impacts of introducing the FI?

Recommendation

To evaluate the impact of the FI, we recommend a **2-armed randomised controlled trial** (RCT) comparing TTF with the FI (treatment) to TTF without FI (control). We recommend an RCT because it offers the most robust evaluation possible, whilst also offering the highest possibility of detecting a significant effect, that is, it is extremely likely to have higher statistical power than alternative quasi-experimental evaluation designs. Despite the risk that the effect of the FI on teacher recruitment and retention is smaller than the minimum detectable effect size (MDES), given the sample size, we believe that an RCT would offer the best chance of measuring the impact of this policy.

The evaluation would work as follows:

- To be eligible for the FI, further education providers would need to register for the trial.
- Providers would then be randomly allocated to either the 'treatment' or the 'control' group. Providers in the **treatment group** would be able to offer the FI to recruits in the FI subjects and would receive an assurance from DfE regarding the number of

FI places they can offer. This could allow providers to advertise the availability of the FI, subject to the inclusion of caveats around eligibility, which is not routinely done due to the lack of certainty around access to the FI.

- Advertising the availability of the FI would be needed for the policy to have any measurable impact on recruitment. Providers in the **control group** would not be able to offer the FI to recruits in the FI subjects, although they could still recruit to these subjects and the core TTF offer would still apply.

The difference between treatment and control groups would be whether they can offer the FI to recruits in the FI subjects; all providers would be able to offer the core TTF programme. The random allocation of providers to treatment and control groups would allow us to infer that any differences in recruitment and retention are likely to be caused by the FI.

The three research questions would be answered by comparing the treatment and control group on:

1) **the number of recruits (per provider) in the FI subjects** in Cohort 2
2) **the 2-year retention rate of recruits in the FI subjects**, that is, the number who complete the 2 years of the TTF programme
3) **potential unintended outcomes** that have been identified in collaboration with DfE and other key stakeholders (including providers who have registered to TTF), which would require additional primary data collection via surveys, for example, a drop in the perceived quality of recruits in non-FI subjects, or a drop in the satisfaction of recruits in non-FI subjects.

An RCT could also explore secondary recruitment and retention outcomes, for example, longer term retention outcomes, and several mechanisms of the Theory of Change for the FI, which has been developed with DfE and other stakeholders.

The **main risk** is that the minimum detectable effect size is too large to be able to detect an effect (i.e., that a true effect of the FI is smaller than what we can find given the sample size). For recruitment, we would expect the MDES to be an increase of between 54% and 82% in the number of FI-subject recruits per provider, depending on the number of providers who take part in the RCT. For retention, we would expect the MDES to be between 8.8% and 17.2% increase in the 2-year retention rate, depending on the number of providers who take part in the RCT. These MDESs seem large, but are not considered completely unachievable if policy and delivery changes were implemented. In addition, our estimates of the MDES are conservative because they do not account for the potential inclusion of covariates in analyses. Adding covariates would likely allow us to detect slightly lower effect sizes.

2. Introduction

2.1. Context and background

The Further Education (FE) sector has documented difficulties with recruitment and retention of teaching staff, as many FE colleges report having 'hard-to-fill' vacancies[1] and high levels of churn amongst new starters.[2] Taking Teaching Further (TTF) is a programme designed to address these issues by providing support for FE providers to recruit and train those with relevant knowledge and experience as FE teachers. Alongside this programme, The Department for Education (DfE) also introduced a financial incentive (FI) pilot of £6,000 paid (over a two-year period) to eligible recruits in some of the most hard-to-fill subject areas. DfE commissioned Verian to conduct an impact evaluation feasibility study to examine options, and to recommend an optimal option, for a quantitative evaluation of the effect of the FI on recruitment and retention.

2.2. The TTF programme

TTF is a programme that supports FE providers to recruit those with relevant knowledge and industry experience to retrain as FE teachers. Since its inception in 2018, the programme has supported the recruitment and training of around 1,000 business and industry professionals across a wide variety of subjects (see Taking Teaching Further programme - GOV.UK (www.gov.uk). During this period, the programme has been extended to include 15 technical routes (aligned to T Levels and the Institute for Apprenticeships and Technical Education approved technical qualifications), English and Maths (introduced in Round 5), and special education needs and disabilities (SEND) specialist teaching roles (introduced in Round 5). The programme was previously administered by the Education and Training Foundation (ETF) but has been directly delivered by DfE since Year 2 of Round 5 (2023).

As part of the TTF programme, FE providers receive up to £18,200 (full offer) or £11,100 (reduced offer) per recruit (subject to eligibility). The full offer covers the cost of:

- Teacher training (to a minimum standard of a Level 5 Diploma in Education and Training)
- 144 hours of intensive support (e.g., mentoring)
- 140 hours (split evenly across Years 1 and 2) of teaching cover to allow the recruit a reduced timetable

[1] Association of Colleges (AoC) College Workforce Survey Summary of findings - 2018/19, AoC. https://d4hfzltwt4wv7.cloudfront.net/uploads/files/AoC-college-workforce-survey-summary-of-findings-2018-19-report-published-2020.pdf

[2] College Staff Survey 2019 follow-up, DfE. https://assets.publishing.service.gov.uk/media/61703f96e90e071979dfecde/CSS_follow_up_survey_Sep_2020.pdf

To access this funding, FE providers submit a recruit registration form on behalf of eligible recruits. FE providers can access funding for up to three recruits initially, although they can register their need for additional places which may become available, subject to demand across the entire programme. Notably, FE providers are only able to apply for the TTF programme on behalf of new recruits after they have been formally appointed to a teaching position. For Round 6, providers can register until 30 November 2023 and register recruits until 31 January 2024.[3]

2.3. The Financial Incentive

For Round 6, which commenced in April 2023, the DfE introduced a non-consolidated financial incentive (FI) pilot of £6,000 offered to eligible recruits in some of the most hard-to-fill subject areas, namely: digital, construction and the built environment, engineering and manufacturing, and maths. This incentive was designed to aid the recruitment and then the retention of new recruits in these subject areas, in part by reducing the salary difference between their existing employment opportunities and a FE teaching role, which was identified as a specific barrier to entry in these subjects.[4] To support the retention of recruits, this incentive is paid in two parts over the course of the programme.

- £3,000 awarded to the recruit upon completion of the first year of the TTF programme (March 2024)
- £3,000 awarded to the recruit upon completion of the second year of the TTF programme (March 2025)

The FI is available on a first-come, first-served basis with a limited number of places available. The recruitment window for the current round closes on 31 January 2024 (in line with the TTF recruit registration deadline). Like the broader TTF programme, providers and recruits are only able to apply for the FI after they have been formally appointed to a teaching position and have eligibility for TTF confirmed. As a result, providers tend not to mention the FI in their advertising materials, severely limiting the ability of the FI to have an impact on recruitment (as the FI will not have played a role in recruiting eligible teachers).

2.4. Aims and objectives of the feasibility study

The aims of this feasibility study were to conduct a review of the FI policy for the purposes of an impact evaluation and determine the most appropriate evaluation approach for an assessment of the impact of the FI for Cohort 2 (FY 24-26), subject to the continuation of the TTF programme and the FI.

[3] https://www.gov.uk/guidance/taking-teaching-further-programme#financial-incentive-for-ttf-recruits
[4] College Staff Survey 2019 follow-up, DfE. https://assets.publishing.service.gov.uk/media/61703f96e90e071979dfecde/CSS_follow_up_survey_Sep_2020.pdf

As part of this feasibility study, this research includes recommendations for the most appropriate evaluation approach, and the necessary sample size, data sources and any necessary adaptations to policy.

2.5. Research questions for the impact evaluation

Any impact evaluation conducted on the FI component of the TTF programme should answer three key research questions (these research questions were discussed and agreed with DfE):

1) Does the introduction of the FI significantly increase teacher recruitment in FE settings?
2) Does the introduction of the FI significantly increase teacher retention in FE settings?
3) What, if any, are the unintended outcomes and negative impacts of introducing the FI?

2.6. Methods

To meet these aims and objectives, this research conducted stakeholder interviews with providers at FE colleges, Independent Training Providers (ITPs), and with the ETF. These discussions were used to assess the validity of the Theory of Change developed in partnership with DfE (this is provided in Section 3.1. with a diagrammatic version in Appendix A). These discussions also provided assurances regarding the quality and availability of survey data which may be used to conduct an impact assessment.

To determine the feasibility of different methodologies, power simulations were conducted to determine the minimum detectable effect size (MDES), given estimates of likely sample size (see Appendix B for full details of the methods used in the power simulations).

3. Outcomes and measurement

To assess the impact of the FI on recruitment and retention, and to quantify the extent of any impact, we need to identify outcome measures for recruitment and retention. The **primary outcomes** are the main focus of the statistical analysis. They will be the basis for conclusions about the effects of the interventions under investigation and were used in power simulations (in order to determine what effect size we would be able to detect given the available sample size). Other **secondary outcomes** may be helpful for explaining or contextualising the main findings, or may measure intermediate mechanisms that have been identified in the Theory of Change, or any unintended consequences (which have been identified in collaboration with DfE and key stakeholders including ETF and providers who have registered to TTF).

Outcomes can be measured at different levels, for example, individual level, subject level, provider level and programme level. The outcomes and outcome measures we recommend are at the lowest level that is feasible, this being recruitment at the level of individual providers and retention at the level of individual recruits. This offers the best opportunity for detecting any impacts (as the relevant sample sizes are largest) and because it allows for exploratory/secondary analyses. For example, if outcomes are only measured at the programme level, such as for the total number of recruits, then it would not be possible to explore whether impacts differ by region and/or demographic or other characteristics.

We selected primary and secondary outcomes in collaboration with DfE, based on their relevance to the research questions, how attributable they are to the intervention, (that is, not too far along the chain of cause and effect in the ToC), and how possible they are to measure (through existing data and/or primary data collection).

3.1. Theory of Change for the FI

The Theory of Change (ToC) details the inputs, the activities, the short/medium/longer term outcomes, and the impacts. A ToC for the TTF programme was published in 2022.[5] We built on this to develop a ToC for the FI, in collaboration with DfE, ETF and several providers who have registered to TTF (three FE colleges and one ITP).

The components of the ToC are as follows (a diagrammatic version is provided in Appendix A):

Inputs (processes):

- College and FE provider time and resource to complete application process

[5] https://www.gov.uk/government/publications/taking-teaching-further-process-evaluation-and-feasibility-study

- Funding for new teacher recruits to supplement salary by around 10%
- DfE time and resource in administering application process
- DfE time and resource in monitoring progress

Activities (processes):

- Recruits with relevant knowledge and industry experience are recruited by colleges and FE providers in FI subjects
- Recruits with relevant knowledge and industry experience undergo initial teacher training and receive early career support
- Recruits receive £6,000, with £3,000 in first year of the programme and £3,000 in second year of the programme

Short-term outcomes:

- Increased number of applicants with relevant knowledge and industry experience
- Previously hard-to-fill vacancies filled
- Teachers have sufficient time and support to complete their initial teacher training
- Teachers do not use their own savings/accrue debt to complete teacher training
- Teachers consider FE teaching ‘on a par’ with working in industry as the FI softens any salary difference with industry

Medium-term outcomes:

- Providers have reduced costs associated with unfilled vacancies (in relation to the cost of supply teachers and overtime)
- New recruits share knowledge gained through recent industry experience with other staff
- More teachers complete their training as part of the programme
- Teachers are happier with teaching as a career
- Learners have fewer lessons taught by supply teachers/ have a more consistent teacher
- Learners are exposed to more up-to-date industry knowledge and feel learning is more relevant

Longer-term outcomes:

- Reduced costs of recruitment because of better retention rates
- Providers are better equipped to deliver technical courses, such as T-levels and technical apprenticeships
- Teachers remain in the profession
- Learners have a higher achievement rate
- Learners have better knowledge about the range of opportunities open to them
- Employers consider that learners going into industry are better equipped to do their jobs

Impacts:

- Reduction in skills gaps in FE to allow FE providers to deliver teaching in some of the most hard-to-fill subject areas
- Greater desirability of teaching in FE among industry professionals
- Learners have better employment outcomes
- Better skilled industry workforce (in key sectors of the economy)

3.2. Recruitment

As the **primary outcome** for recruitment, we recommend using the number of recruits, since this is likely to be the most sensitive measure of the impact on recruitment. That is, most likely to detect an effect of the FI on recruitment, conditional on there being an effect of the FI on recruitment. For example, alternative measures, such as the number of vacancies in FI subjects, would likely to be less sensitive to detect the effect of the FI because vacancies filled by TTF recruits make up just a proportion of the total number of vacancies filled.

For measuring the impact of the FI, we recommend using count data at the level of individual providers (the lowest level that is feasible for the evaluation of the programme across all FI subjects) and that this uses the number of TTF recruits in Round 7 (Cohort 2 of the FI) in the FI subjects, as this is likely to be the most sensitive measure of the impact on recruit numbers. For example, if we were to use recruit numbers for the whole TTF programme, the measure is likely to be less sensitive as not all subjects receive the FI.

This primary outcome measure would use internal DfE data (TTF programme recruitment records).

As **secondary outcome** measures for recruitment we recommend that an evaluation also looks at the following:

- Rate of unfilled vacancies at the end of the academic year in FI subjects (at the provider level for colleges/providers in the TTF programme): This would investigate the impact on vacancy rates and could use the Further Education Workforce Data Collection (FEWDC) and/or additional data collection.

- Providers' perception of the quality of TTF recruits in FI subjects (on a Likert scale, one per recruit): This would investigate impact on the quality of those recruited and would require additional data collection.

3.3. Retention

As the **primary outcome** for retention, we recommend using the proportion of recruits still in post at the end of their second year, that is, at the end of the 2 years of the TTF programme. This is most likely to capture the full effect of the FI, which runs for the 2 years of the TTF programme and is therefore likely to be the most sensitive measure, that is, it is most likely to detect an effect of the FI on retention, conditional on there being an effect of the FI on retention. The disadvantage, compared to shorter time periods, is that it will take longer for the outcome measure to become available.

For measuring the impact of the FI, we recommend using data at the level of individual recruits (the lowest level possible) and using only the FI subjects, as this is likely to be the most sensitive measure of the impact on retention numbers. For example, if we were to use retention for the whole TTF programme, the measure is likely to be less sensitive because not all subjects receive the FI.

This primary outcome measure would use internal DfE data (TTF programme records on retention) and in future FEWDC.

We also recommend conducting additional primary data collection as part of an evaluation. Providers would be asked to submit specific information on retention at regular intervals, including them specifying when exactly a recruit left their post, even if this was after all TTF claims were paid. This could be done at the same time as collecting other data from providers.

As **secondary outcome** measures for retention we recommend that an evaluation also looks at the following:

- Proportion of FI-subject recruits still in post at the end of 1 year, using data at the recruit level: Compared to 2-year retention, this has the advantage that the data

would be available sooner, but it would not be expected to capture the full effect of the FI (as the FI is given over the 2 years of the TTF programme). This would use internal DfE data (TTF programme records on retention). As for 2-year retention, we recommend additional primary data collection in the form of providers being asked to submit specific information on retention at regular intervals.

- Two separate measures for the proportion of FI-subject recruits still in post at the end of 3 and 5 years (1 and 3 years after the programme finishes), using data at the recruit level: These would capture longer term outcomes, but it would take longer for the data to be available. This would use internal DfE data (TTF programme records on recruitment) matched to the FEWDC. The main limitation of this is that the FEWDC currently has imperfect compliance by providers in completing this information. As for 2-year retention, we recommend additional primary data collection in the form of providers being asked to submit specific information on retention at regular intervals, but this is also likely to have issues with compliance (especially over longer time periods).

- Mean length of time FI-subject recruits stay in post, using data at the recruit level: This would capture longer term outcomes but it would take longer for the data to be available and there is the limitation that this measure is conditional on having left. For example, after 10 years, a measure could only say how long those that left within 10 years stayed on average but nothing about those that are still in post after 10 years and how long these might stay (which may well differ by groups that are being compared). This would share the same limitations and has the same recommendations to mitigate these limitations as 3/5-year retention above.

3.4. Intermediate mechanisms of the Theory of Change

We recommend that an evaluation also looks at several secondary outcome measures to explore intermediate mechanisms of the ToC, although these would all require primary data collection. Interviews with providers suggest that providers are willing to complete such additional data collection. In addition, providers believe that recruits would be willing to complete additional data collection in the form of questionnaires, but they recommend that these are administered via the providers and completed during TTF training time to maximise completion rates.

Primary data collection via providers who are asked to complete this information at regular intervals via questionnaires (this could be done at the same time as collecting other data outlined above):

- Number of applications (from those with suitable knowledge/experience) per TTF job advert in the FI subjects

- Unfilled vacancy costs in the FI subjects

- Recruitment costs in the FI subjects

Primary data collection via recruits who are asked to complete this information at regular intervals via questionnaires:

- Happiness of recruits in the FI subjects (on a Likert scale)
- To what extent recruits in the FI subjects consider teaching to be ‘on a par’ with working in industry (on a Likert scale)
- Other indicators related to recruits, for example, job satisfaction and confidence could also be looked at (on a Likert scale)

3.5. Unintended outcomes

Several potential unintended outcomes/consequences have been identified in collaboration with DfE and other key stakeholders (including ETF and providers who have registered to TTF). As secondary outcome measures to explore and monitor these, we recommend that an evaluation also looks at the following:

- Total number of TTF recruits during the round in the non-FI subjects (count data at the level of individual providers): This would look at a potential unintended outcome where potential recruits (who are qualified for multiple subjects) ‘switch’ from non-FI subjects to FI subjects when the FI is available. This measure would use internal DfE data (TTF programme recruitment records).
- Providers’ perception of the quality of TTF recruits in non-FI subjects (on a Likert scale, 1 per recruit): This would look at a potential unintended outcome where non-FI subjects attract lower quality recruits when the FI is available. This could be due to, for example, lower competition in these subjects when there is the FI in FI subjects. This would require additional data collection and could be collected at the same time as other data is collected from providers.
- In the year after the FI ends (Year 3 after being recruited), happiness of recruits in the FI subjects and to what extent recruits in the FI subjects consider teaching to be ‘on a par’ with working in industry (using Likert scales): This would investigate a (potential) drop-off in satisfaction after the FI ends. This would require additional data collection from recruits (via questionnaire) administered at regular time points.
- Using a base of those still in post after 2 years, what proportion of these FI-subject recruits are still in post at the end of 3 years: This would look at a (potential) drop-off in retention after the FI ends. This would use the same data as 3-year retention discussed above (and would have the same limitations).

- Happiness of recruits in the non-FI subjects and to what extent recruits in the non-FI subjects consider teaching to be ‘on a par’ with working in industry (using Likert scales): This would look at potential unintended outcomes where other members of staff that do not receive the FI might feel undervalued in some way. This would require additional data collection from recruits (via questionnaire).

- If not all providers offer the FI, to what extent recruits in FI subjects considered different providers (on a Likert scale): In combination with other data (on recruitment), this can explore if there is evidence for a shift in recruitment from providers not offering FI to those that are (if there is then we would expect a positive correlation between considering different providers and going to a provider offering the FI). This would require additional data collection from recruits (via questionnaire).

- To what extent recruits in the FI subjects felt that they wanted to make sure they got a job with a provider sooner, rather than later: This would look at whether there is any evidence for a change in the timing of the decision to move into teaching in the FE sector, which might occur instead of a change in whether someone moves into teaching or not. For example, there could be an increase in applicants/recruits as they consider there to be a risk that the policy will not continue indefinitely but these people would have applied at some point in the future. This would require additional data collection from recruits (via questionnaire).

- If a cap on FI places is reached (despite an evaluation being designed in such a way that this is unlikely), 2-year retention rates (as outlined above) for those in FI subjects and recruited by providers who were able to offer FI places but did not receive the FI as the cap on places was reached: This would look at a potential negative outcome related to different recruits at the same provider (and in the same role) differing in whether they get the FI or not. This would use the same data as 2-year retention discussed above (and would have the same limitations).

- Giving the FI to people who would have been recruited and stayed for the long term anyway (without the FI): To some extent this is unavoidable, and the recommended evaluation will be able to explore this using the primary outcome measures of recruitment/retention and robust comparisons to a comparison group. This will be discussed further in Chapters 4 and 5.

4. Impact evaluation options considered

Impact evaluations aim to assess what changes have occurred and the scale of those changes, by comparing outcomes in the 'treatment group' to those of a counterfactual 'control group' that did not receive the intervention.[6] Randomised controlled trials (RCTs) are often considered to be the gold standard because random assignment of 'units' to treatment and control groups allows a robust inference that the intervention caused any differences in outcomes. However, other options for constructing a counterfactual comparison group may be used when RCTs are not feasible. We also considered regression discontinuity design (RDD), difference in difference (DiD), and synthetic difference in difference (SDiD).

We recommend using an RCT to assess the impact of the FI. An RCT should be feasible (subject to policy changes and an increase in the number of recruits achieved) and it is highly likely to have higher power to detect an effect than any other approach.

This chapter will outline the details of how each of the designs that we considered could be used to evaluate the FI, and each design's strengths and weaknesses. Chapter 5 will detail how our recommendation of an RCT answers the impact evaluation research questions from Section 2.5. and can also be used to explore mechanisms of the Theory of Change.

4.1. Randomised controlled trial (RCT)

RCTs involve a comparison between 2 or more groups (at least 1 control group that is not given the intervention and 1 treatment group that is given the intervention), with random assignment into these groups. This random assignment strives to ensure that there are no differences (known or unknown) between groups and any differences in outcome measures between the groups can be reliably attributed to the intervention.

Randomisation

For an RCT, we recommend that providers are the 'unit of randomisation', which is allocated to either the treatment or control group. This means that providers will know prior to recruitment whether they can offer FI places and how many of these they can offer (which is not currently the case). Therefore, the FI can be explicitly mentioned in the marketing of vacancies in the treatment group, subject to the inclusion of caveats around eligibility, which stakeholder interviews have highlighted is rarely happening under the current mode of delivery. This will allow the FI to impact on recruitment in the manner envisaged in the Theory of Change. Although simple random sampling can be used, to avoid imbalance in the groups we recommend stratified randomisation using characteristics such as whether the provider is an ITP, provider region, whether a

[6] https://www.gov.uk/government/publications/the-magenta-book

provider is in a rural or urban location, and whether the provider is targeted by any other policies that aim to increase teacher recruitment/retention.

An alternative would be to conduct the randomisation at the level of individual places. Although it would be possible for individual places to be randomised in this way, in order for the FI to impact recruitment, providers would need to know (in advance of advertising) which individual vacancies have the FI and which do not. This would result in at least some providers advertising FI and non-FI vacancies at the same time in the same subject. In addition to complicating the recruitment/advertising process, this has a high risk of causing bias (and therefore impacting the validity of any results) through applicants switching from a non-FI vacancy to an FI vacancy with the same provider (when they would have applied and been recruited to the non-FI vacancy in the absence of the FI vacancy). Although a similar concern exists with randomisation across providers (that recruits switch to providers who offer the FI), this risk is more likely with randomisation of individual places.

Design outline

A 2-armed trial would have the treatment group as providers (and the teachers recruited by these providers) who can offer the FI in the FI subjects and the control group as providers (and the teachers recruited by these providers) who cannot offer the FI in the FI subjects. The full outline of how we recommend such an RCT would work in practice is provided in Section 5.1.

Power simulations

We conducted power simulations to estimate the effect sizes that the FI would need to have for significant effects to be likely to be detected, given different sample sizes. (See Appendix B for full methodological details.) Below are the smallest increases in recruitment and retention that we would be able to detect, given the sample size, or the 'minimum detectable effect size'. Note that we would be able to detect smaller effect sizes for both primary outcomes if the number of subjects covered by the FI – and the number of FI recruits – were increased.

For **recruitment**, we would expect the minimum detectable effect size to be an increase of between 54% and 82% in the number of FI-subject recruits per provider, depending on the number of providers who take part in the RCT (Table 1). Based on previous rounds of TTF data, if 150 providers (in total, across both arms) take part in the trial, then there would need to be an 82% increase in FI-subject recruits (on average) per provider to have an 80% probability of detecting this effect on recruitment. If 200 providers take part in total, then there would need to be an 68% increase in FI subject recruits (on average) to have an 80% probability of detecting this effect. If 300 providers take part in total, there would need to be an 54% increase in FI subject recruits (on average) to have an 80% probability of detecting this effect.

Table 1: Recruitment effect sizes for a 2-armed RCT at different sample sizes (number of providers), sample size is the total of both arms.

Number of providers	Mean count of FI-subject recruits per provider in the control group	Minimum detectable effect size (MDES): Mean minimum increase in the count of FI-subject recruits per provider from control to treatment (% change) in order to be able to detect a statistically significant difference	Minimum mean count of FI-subject recruits per provider in the treatment group, in order to be able to detect a statistically significant difference
150	1.13	0.93 (82%)	2.06
200	1.13	0.77 (68%)	1.90
300	1.13	0.61 (54%)	1.74

For **retention**, we would expect the minimum detectable effect size to be an increase in the 2-year retention rate of between 8.8% and 17.2% (percentage points), depending on the number of providers who take part in the RCT (Table 2). The mean probability of an FI-subject recruit still being in post after 2 years is currently 73%, taking the average from Rounds 1-4. If, for example, there are 150 FI-subject recruits in total across both arms, there would need to be a 17.2 percentage point increase (to 90.2%), to have an 80% probability of detecting the effect of the FI on retention. If there are 300 FI-subject recruits in total, then a 12.9 percentage point increase (to 85.9%) would be needed to have an 80% probability of detecting this effect. For 700 FI-subject recruits in total, an 8.8 percentage point increase (to 81.8%), would be needed to have an 80% probability of detecting this effect.

Table 2: Retention effect sizes for a 2-armed RCT at different sample sizes (number of recruits in the FI subjects), sample size is the total of both arms.

Number of recruits in the FI subjects	Mean probability of an FI-subject recruit still being in post after two years in the control group	Minimum detectable effect size (MDES): Mean minimum (percentage point) increase in the probability of an FI-subject recruit still being in post after two years from control to treatment	Minimum mean probability of an FI-subject recruit still being in post after two years in the treatment group, in order to be able to detect a statistically significant difference
150	73%	17.2%	90.2%
600	73%	12.9%	85.9%
700	73%	8.8%	81.8%

Strengths

The key methodological strength of this design is that RCTs randomise units (such as recruits/providers) to treatment and control groups. Randomisation strives to ensure that control and treatment groups are balanced on observed and unobserved characteristics. This, in turn, increases the probability that the difference between the control and intervention group is due to the impact of the intervention,[7] thereby allowing for causal inference. In addition, it is the most likely of any of the designs considered to detect an effect related to the FI, that is, it is likely to have the highest power. The minimum detectable effect sizes reported by the simulations seem large, but are not considered completely unachievable if policy and delivery changes were implemented, especially given that these power simulations are conservative as they do not model the inclusion of any potential covariates which would likely improve power slightly. Potential covariates that could be included are estimates from providers (collected before randomisation into treatment and control, for example, at the time of provider registration) for how many recruits they expect in the FI subjects and the mean salary they expect to be able to offer recruits in the FI subjects. Power would also be increased if the number of subjects covered by the FI – and the number of FI recruits – is increased in future rounds.

[7] https://www.ukri.org/publications/evaluation-framework/

Weaknesses

Considering the RCT from a policy perspective, some providers would need to be randomly allocated to the control group (who cannot offer the FI), and all providers would need to agree to the possibility of this happening in advance; however, this would be a condition of being able to offer the FI, so we do not anticipate this being a problem. Further, in the first cohort of the FI, funding was never guaranteed for any providers, so providers might welcome the certainty of the funding and ability to advertise, with the inclusion of caveats around eligibility, during recruitment.

One weakness of this RCT design is that any analysis of the retention outcomes of recruits uses only those people who are recruited, so there is a risk that (if the FI has an effect on recruitment), then the treatment and control groups will no longer be comparable for the analysis of retention. However, the trial as a whole is comparable to how the FI policy would likely be implemented outside of a trial setting, that is, potential recruits would know whether they will receive the FI. Therefore, the groups should closely represent what would happen in the absence of a trial and give an estimate of the impact of the FI on retention (taking into account any changes in recruit characteristics because of the FI). In addition, the potential that the intervention and control groups will not be comparable can be mitigated by matching groups on characteristics responsible for observed imbalance and/or incorporating control variables, for example, age and gender, into the analysis. Alternatively, a composite measure of recruitment and retention at the provider level, for example, the number of recruits who are recruited and are still in post after two years, per provider (where the randomisation still holds), could be used. However, this would not allow a separation of the effects of recruitment from retention and is therefore not recommended.

4.2. Regression discontinuity design (RDD)

An RDD assigns units (providers in this case) to treatment and control groups based on whether they are above or below a threshold on a pre-selected variable (the eligibility index). Units that are close to the threshold are expected to be randomly distributed with respect to other on known and unknown variables, and their outcomes are compared. The treatment and control groups would be the same as in the 2-armed RCT (providers in the treatment group can offer the FI and those in the control group cannot).

Allocation

Any pre-selected variable can be used to allocate providers to treatment or control group in an RDD. To maximise feasibility from a policy perspective we would recommend that the pre-selected variable has some policy relevance, for example, past vacancy rates at the provider level. This would be used to set a threshold vacancy rate, with providers allocated to the treatment group if their vacancy rate was above the threshold and to the control group if their vacancy rate was below the threshold.

Design outline

The full outline of how such a design would work in practice is as follows:

1. Providers register for TTF as normal but know that they are not guaranteed to be able to offer FI funded places.

2. Providers would be allocated to either a treatment or a control group using an eligibility index. The index could be, for example, the most recent provider level vacancy rates available. This would use the Further Education Workforce Data Collection (FEWDC) for the 22/23 academic year for which collection is scheduled to close in January 2024.[8] Providers would be allocated to the treatment group if their vacancy rates were above the set threshold (in the most recent academic year that the data is available) and to the control group if their vacancy rates were below the threshold. The threshold vacancy rate would be set such that the resulting treatment and control groups would contain roughly equal numbers of providers (although this is not crucial to the design). Any providers with missing data on the selection variable would have to be omitted from the analysis and, due to the limited number of FI places, we would recommend that they do not receive the FI.

   Treatment: Providers can offer the FI in the FI subjects, this being in addition to the core TTF programme/offer (providers receiving the usual funding).

   - When providers are allocated to the treatment group, they receive assurance that the FI is available for a set number of recruits (subject to them meeting the eligibility criteria) in the FI subjects (allowing them to advertise on the basis of this, with the inclusion of caveat around eligibility). This number would be set in a way such that it would be unlikely for the cap on places to be reached.
   - If providers have confirmed recruits for the assured places but there is still funding available (that has not been provisionally allocated to assured places), they can request additional FI places on a first-come, first-served basis (one at a time with providers only being allowed to request more once an additional recruitment is confirmed). Providers are informed that funding for each additional place is assured, subject to recruits meeting the eligibility criteria (allowing them to advertise, with the inclusion of caveats around eligibility, on the basis of this).

   Control: Providers cannot offer the FI in the FI subjects, but the core TTF programme/offer is unchanged. Thus, they can still recruit in the FI subjects through TTF (and the providers will receive the core TTF funding) but recruits will not receive the FI.

[8] https://www.gov.uk/government/publications/further-education-workforce-data-collection

3. Monitor/collect outcome measures for both recruitment (using provider level data) and retention (using recruit level data) as outlined in Chapter 3.

Strengths

A strength of this design is that the variable that is used for allocation to treatment and control groups can be chosen so that it that has some justification from a policy perspective, for example, provider level vacancy rates (as opposed to random assignment in an RCT). As with an RCT, some providers would need to be allocated to not being able to offer the FI, but we do not anticipate this being a problem (and even less so if a policy relevant selection variable is used and referenced when providers are informed of their allocation). In the first year of the FI, funding was never guaranteed to any provider, so they might welcome certainty of the funding and ability to advertise, subject to the inclusion of caveats around eligibility, during recruitment.

Weaknesses

As with an RCT any analysis of the retention outcomes of recruits uses only those people who are recruited, so there is a risk that (if the FI has an effect on recruitment), then the treatment and control groups will no longer be comparable for the analysis of retention. As with the RCT design, this is not a major concern in terms of the validity of any conclusions and this risk can be mitigated.

A major weakness is that (for the same number of providers/recruits) an RDD is likely to have a significantly lower probability of detecting an effect compared to an RCT (that is, it has significantly lower power). In addition, an RDD is generally seen as less robust than an RCT, the main reason being that any conclusions may not apply to those providers (and their recruits) that are further from the threshold.

As outlined above, it would be recommended to use a policy relevant selection variable but if this data is not available for some providers, then those providers (and their recruits) would need to be excluded from the analysis. This would reduce the sample size and reduce the likelihood of detecting an effect (that is, it would lower power). In addition, data for the selection variable would have to be available sufficiently early to allow for the allocation of providers to treatment and control groups before they find out the outcome of their TTF provider registration.

4.3. Difference in difference (DiD)

This measures changes in time trends of an outcome of interest, before and after the intervention, comparing a group that is given the intervention, to a control group that is not subject to the intervention. For example, one outcome of interest is the number of recruits (in a particular set of subjects), the intervention is the introduction of the FI in

Round 6 (Cohort 1 of the FI), the group that is given the intervention is the FI subjects and the control group is the non-FI subjects.

Design outline

As outlined in Section 4.1., the implementation of the FI in Round 6 (Cohort 1 of the FI) has not allowed the FI to impact recruitment in the way that it would in a wider implementation (as providers were not able to guarantee to potential recruits that FI funding was still available, so did not routinely advertise the FI in recruitment materials). Therefore, we would recommend that a DiD design would change the delivery of the FI so that, from Round 7 (Cohort 2 of the FI), FI funded places can be guaranteed for specific roles and mentioned in the recruitment materials, subject to the inclusion of caveats around eligibility. This would be the same as the treatment groups in the RCT and RDD designs, except that all providers would be in this group. The full outline of how such a design would work in practice is as follows:

1. Providers register for TTF as normal but know that they will be guaranteed to be able to offer a certain number of FI funded places which will be confirmed before the majority of recruitment starts.

2. Providers receive assurance that the FI is available for a set number of recruits (subject to recruits meeting the eligibility criteria) in the FI subjects (allowing them to advertise, with the inclusion of caveats around eligibility, on the basis of this). This number will be set in a way such that it is unlikely to cause the cap on places to be reached.

    - If providers have confirmed recruits for the assured places but there is still funding available (that has not been provisionally allocated to assured places), they can request additional FI places on a first-come, first-served basis (one at a time with providers only being allowed to request more once an additional recruit is confirmed). Providers are informed that funding for each additional place, subject to a recruit meeting the eligibility criteria, is assured (allowing them to advertise this, with the inclusion of caveats around eligibility).

3. Monitor/collect outcome measures for both recruitment (using provider level data) and retention (using recruit level data) as outlined in Chapter 3 (for both FI subjects and non-FI subjects).

Strengths

A strength of this design from a policy perspective is that no provider/recruit will be allocated into a control group where they are not able to offer FI funding. There is therefore less change from business as usual/how the policy would likely be delivered in the longer term.

Another strength is that an analysis can be done using data from Round 6. However, such an analysis is unlikely to be able to detect any significant effect (and certainly less likely than an RCT) as the effect sizes (for the same number of recruits) would need to be greater than those reported in Table 2 in Section 4.1. In addition, such an analysis would only be able to look at retention outcomes and not recruitment since advertising of the FI at the recruitment stage has not been possible previously. A design that modifies the delivery of the FI for Round 7, as outlined above, would be able to look at recruitment.

Weaknesses

A major weakness is that (for the same number of providers/recruits) a DiD design is likely to have a significantly lower probability of detecting an effect compared to an RCT, (that is, it is likely to have significantly lower power). In addition, DiD is generally seen as less robust than an RCT. One reason for this is that it assumes that the outcome variable moves in parallel (for the different groups) before the intervention and would have continued to do so had the intervention not occurred (the parallel trends assumption).

4.4. Synthetic difference in difference (SDiD)

SDiD uses historical data to create a synthetic control that matches pre-treatment outcomes of the intervention group. For example, synthetic controls could be created by combining multiple non-FI subjects that are similar to the FI subjects in their number of recruits. The approach is similar to a DiD model in that the FI effect would be estimated by comparing the change in outcomes between the FI subjects and the synthetic control subjects before and after the FI is introduced.

Design outline

An SDiD would have the same design as DiD from a practical perspective. The only difference is in the analysis.

Strengths

SDiD shares the same strengths as DiD and has the added strength that it can be more robust than DiD as it relaxes the assumption that the outcome variable moves in parallel (for the different groups) before the intervention.

Weaknesses

This design shares the same limitation as DiD regarding the analysis of recruitment using Round 6 data. We would therefore recommend that a SDiD design modifies delivery of the FI from Round 7 so that FI funded places can be guaranteed for specific roles (conditional on recruits meeting the eligibility criteria) and mentioned in the recruitment materials.

As with DiD, a major weakness is that (for the same number of providers/recruits) an SDiD design is likely to have a significantly lower probability of detecting an effect compared to an RCT (that is, it is likely to have significantly lower power). In addition, SDiD is generally seen as less robust than an RCT. Another weakness is that, for good estimation, SDiD requires more pre-intervention (historical) data than DiD.

5. Recommendation

We recommend a 2-armed RCT comparing TTF with the FI (treatment) to TTF without FI (control) as described in the Section 4.1. We recommend this as this design offers the most robust evaluation possible whilst also offering the highest possibility of detecting a significant difference (that is, it is likely to have higher statistical power than any of the alterative evaluation designs). To note, the suggested delivery changes are required for the recommendation to be possible.

5.1. Design details

As outlined in Section 4.1., we recommend a design that would allow providers to know, prior to recruitment during the round, if they could offer FI places and how many TTF places with the FI they can offer (which is not currently the case):

1. Providers register on to the trial as part of the current process for registering for TTF. To maximise the number of providers and recruits in the trial, we would recommend the trial being the only mechanism for recruits to be able to receive the FI.

2. Providers are randomly allocated to either a treatment or a control group and informed of this.

   Treatment: Providers can offer the FI in the FI subjects, this being in addition to the core TTF programme/offer (providers receive the core TTF funding).

   - When providers are allocated to the treatment group, they receive assurance that the FI is available for a set number of recruits (subject to these recruits meeting the eligibility criteria) in the FI subjects (allowing them to advertise, with the inclusion of caveats around eligibility, on the basis of this). This number will be set in a way such that it is unlikely to cause the cap on places to be reached.
   - If providers have confirmed recruits for the assured places but there is still funding available (that has not been provisionally allocated to assured places), they can request additional FI places on a first-come, first-served basis (one at a time with providers only being allowed to request more once an additional recruitment is confirmed). Providers are informed that funding for each additional place, subject to each recruit meeting the eligibility criteria, is assured (allowing them to advertise, with the inclusion of caveats around eligibility, on the basis of this).

   Control: Providers cannot offer the FI in the FI subjects, but the core TTF programme/offer is unchanged. Thus, they can still recruit in the FI subjects through TTF (and the providers will receive the core TTF funding) but recruits will not receive the FI.

3. Monitor/collect outcome measures for both recruitment (using provider level data) and retention (using recruit level data) as outlined in Chapter 3.

5.2. How the design answers the research questions

RQ1: Does the introduction of the FI significantly increase teacher recruitment in FE settings?

An RCT would show whether the FI significantly impacts recruitment when compared to the 'core' TTF offer (conditional on the sample and effect sizes being at least as large as assumed in the power simulations in Section 4.1.). This would use the primary outcome measure of the number of TTF recruits in Round 7 (Cohort 2 of the FI) in the FI subjects, comparing the treatment group to the control group.

If there is a statistically significant effect, the evaluation would be able to quantify this impact in the number of additional recruits into the FI subjects for Round 7 per provider (this can also be extrapolated to provide a total number of recruits across all providers).

As detailed in Section 3.2., secondary outcomes for recruitment can also be investigated as part of an impact evaluation (such as vacancy rates). In an RCT setting, it would be relatively straightforward to incorporate new primary data collection (via surveys) on some of these secondary outcomes.

RQ2: Does the introduction of the FI significantly increase teacher retention in FE settings?

An RCT would show whether the FI significantly impacts retention when compared to the 'core' TTF offer (conditional on the sample and effect sizes being at least as large as assumed in the power simulations in Section 4.1.). This would use the primary outcome measure of the proportion of recruits (who were recruited in Round 7) still in post at the end of their second year (that is, at the end of the 2 years of the TTF programme) and compare the treatment group with the control group.

If there is a statistically significant effect, an evaluation would be able to quantify this impact as a percentage point increase in 2-year retention in the FI subjects.

As detailed in Section 3.3., secondary outcomes for retention can also be looked at as part of an impact evaluation, including some longer-term outcomes (including, 3-year and 5-year retention).

RQ3: What, if any, are the unintended outcomes and negative impacts of introducing the FI?

An RCT would be able to monitor the (potential) unintended outcomes that have been identified in Section 3.5, for example, for the non-FI subjects, a reduction in recruit

numbers (if some potential recruits in non-FI subjects switch to FI subjects) or a reduction in the quality of those recruited (as there may be less competition in non-FI subjects when there is the FI available in FI subjects). As detailed in Section 3.5., most outcomes would require additional data collection via survey(s), but these are relatively straightforward to carry out as part of an RCT. One reason being that data can be collected as part of the trial registration process and ongoing engagement with the trial. Any analysis would however be secondary, and it is unlikely that significant effects would be found unless the effects of any of the unintended outcomes were relatively large.

5.3. How the design explores the mechanisms of the Theory of Change

As discussed in Section 3.4., we also recommend that an impact evaluation using an RCT explores some of the mechanisms of the Theory of Change. As detailed in Section 3.4, this exploratory analysis would use secondary outcomes, for example, the number of applications that providers receive and the happiness of recruits (as measured using a Likert scale), to compare the treatment group with the control group. All outcomes would, however, require new primary data collection via survey(s) and there is a risk that any analysis is limited by low compliance in completing the surveys. This is of a particular concern in the control group (as they receive no benefits from the FI) and, if low compliance is significant, this may impact the validity of any analysis.

5.4. Risks and mitigations

The **main risk** is that the **minimum detectable effect size** is too large to be able to detect an effect. For recruitment, we would expect the minimum detectable effect size to be a mean increase (between the control and treatment group) of between 54% and 82% in the number of FI-subject recruits, across all providers, depending on the number of providers who take part in the RCT. For retention, we would expect the minimum detectable effect size to be a mean increase in the 2-year retention rate of between 8.8% and 17.2% (percentage points), depending on the number of providers who take part in the RCT. These minimum detectable effect sizes seem large, but are not considered completely unachievable if policy and delivery changes were implemented; especially for retention, and especially given that these power simulations are conservative (because they do not model the inclusion of any potential covariates which would likely improve power slightly). Power simulations assumed no changes in policy, including only the current subjects covered by the FI, but power would be higher if the number of FI subjects is increased in future rounds.

The **main change to policy** is telling providers in advance whether or not they will receive the FI for their vacancies in FI subjects (conditional on recruits meeting the eligibility criteria), which has the possibility for **some reputational risk** because some providers who were able to offer the FI in Round 6 (Cohort 1 of the FI) will not be able to

offer it in Round 7 (Cohort 2 of the FI) (and will know that other providers are able to offer the FI). The change would give providers certainty about the FI, and allow them to advertise it in recruitment materials, subject to the inclusion of caveats around eligibility, allowing the FI to impact recruitment.

There would need to be a **cap on the number of FI places each provider could offer**. We expect to be able to allocate the places such that the cap on places would not be reached.

Appendix A: Taking Teaching Further financial incentive Theory of Change

Situation	Teachers and leaders are one of the biggest determinants of outcomes for learners in Further Education (FE). However, FE has faced difficulties in terms of recruitment and retention – which has been more acute in certain subjects. The FE sector, therefore, needs a sufficient supply of high-quality teachers and leaders with relevant knowledge and industry experience coming into and staying in the sector, to ensure that learners can acquire the outcomes they need for their own prosperity and also for greater national prosperity.	**Aims**	The aim of the financial incentive is to provide targeted support for recruits with relevant experience to teach in some of the most hard-to-fill subject areas where the impacts are already being felt.

Inputs and Activities	**Outputs**			**Impacts**
Inputs College and FE provider time and resource to complete the application process. Financial incentive pilot for teachers in some of the hardest-to-fill subject areas. DfE time and resource in administering application process.	**Short term** Increased number of applicants with relevant knowledge and industry experience Previous hard-to-fill vacancies filled. Teachers have sufficient time and support to complete their initial teacher training	**Medium term** Providers have reduced costs associated with unfilled vacancies (i.e., supply teachers and overtime) New recruits share knowledge gained through recent industry experience with other staff	**Long term** Reduced costs of recruitment because of better retention rates Providers are better equipped to deliver technical courses, such as T Levels and technical apprenticeships Teachers remain in the profession	Reduction in skills gaps in FE to allow FE providers to deliver teaching in the hard-to-fill subjects Greater desirability of teaching in FE among industry professionals Learners have better employment outcomes Better skilled industry workforce (in key sectors of the economy)

DfE time and resource in monitoring progress. **Activities** People with relevant knowledge and industry experience are recruited by colleges and FE providers in financial incentive (FI) subjects* People with relevant knowledge and industry experience undergo initial teacher training and receive early career support. Recruits receive £6,000, with £3,000 in the first year of the programme and £3,000 in the second year of the programme.	Teachers do not use their own savings/accrue debt to complete teacher training Teachers consider FE teaching 'on a par' with working in industry as the FI softens any salary with difference with industry	More teachers complete their training as part of the programme Teachers are happier with teaching as a career Learners have fewer lessons taught by supply teachers/have a more consistent teacher Learners are exposed to more up-to-date industry knowledge and feel learning is more relevant	Learners have a higher achievement rate Learners have better knowledge about the range of opportunities open to them Employers consider that learners going into industry are better equipped to do their jobs	

*FI subject: digital; construction and the built environment, engineering and manufacturing; and maths

Appendix B: Power simulations methodology

The power simulations were run as follows:

1. For both primary outcomes, we first identified the data generating process through which observations for each outcome were generated, and the distribution that we thought was a consequence of this process. For example, for retention the data generating process is whether a recruit stays for 2 years after being recruited or does not. The distribution that is the consequence of this data generating process in a population of recruits is binomial.

2. We used existing TTF data to obtain estimates for parameter values of the distribution, per outcome.

3. We used these values to simulate a distribution of outcome variable observations, per outcome.

4. We then analysed the simulated data using a model that best fit the simulated data (for recruitment a negative binomial model given the negative binomial distribution of the recruitment outcome; for retention a logistic regression model given the binomial distribution of the retention outcome) to obtain an estimate of the intervention effect.

    - The estimate of the intervention effect was then evaluated against the p-value acceptance criterion ($p < 0.05$) to assess whether the simulated study found a significant intervention effect.

5. Stages 3 and 4 were repeated 1,000 times for each specified sample size and effect size combination (72,000 times in total, for the recruitment outcome). We used 1,000 simulations per scenario to ensure high accuracy of our estimates.

6. We then calculated the statistical power (the probability of finding a significant effect if it is truly there) using the number of significant intervention effects as a proportion of the 1,000 simulations per scenario.

    - From all the scenarios we evaluated, we reported (in Tables 1 and 2, Section 4.1) only the estimates for the smallest effect size per sample size that reached the conventional threshold for power.

For the **recruitment** primary outcome (reported in Table 1 in Section 4.1), power estimates were obtained by storing values from simulations involving generation of negative binomial distributions and the use of negative binomial regression models. Below are the assumptions underlying these simulations:

- Significance level = 0.05

- Power = 0.8 (lowest MDES where power is 0.8 or greater)
- Two tailed tests
- Randomisation at the provider level
- Sample size (number of providers) equally split between treatment and control
- Mean count of recruits at the baseline based on Round 5 data for the current four FI-eligible subjects
- The count of recruits follows a negative binomial distribution (based on the distribution of the count of recruits in FI-eligible subjects in Round 5 data)

For the **retention** primary outcome (reported in Table 2 in Section 4.1), power estimates were obtained by storing values from simulations involving generation of binomial distributions and the use of logistic regression models. Below are the assumptions underlying these simulations:

- Significance level = 0.05
- Power = 0.8 (lowest MDES where power is 0.8 or greater)
- Two tailed tests
- Sample size (number of recruits) equally split between treatment and control (note that if recruitment is successful then there will be more in the treatment group than in the control)
- Baseline probability of retention after two years is 73% (this is based on the average retention proportion from Rounds 1 to 4 using internal TTF management information data)
- We assumed that retention was a binary event, whereby a recruit could either leave the post (0) or still be in post after 2 years (1)

Department for Education

Reference: RR1454

ISBN: 978-1-83870-570-1

For any enquiries regarding this publication, contact us at: www.education.gov.uk/contactus

This document is available for download at www.gov.uk/government/publications