

Research protocol

University of Birmingham: Forward Thinking evaluation

| VERSION | DATE | REASON FOR REVISION/NOTES |
|--|----------|--|
| <i>Any changes to the design to be agreed between the implementation partner(s), evaluator and TASO. Note any agreed changes in the table below.</i> | | |
| 3. | Oct 2022 | Study was re-designed due to data access challenges. We could not gain access to linked National Pupil Database data and Higher Education Statistics Agency data for a third party dataset. Therefore, the study has been simplified - as set out in the protocol. |
| 2. | Jan 2021 | Change of TASO staff – Rain Sherlock now TASO lead on MIOM local evaluations. |
| 1. [original] | Nov 2020 | NA |
| Pre-registration | | This design has been pre-registered on The Open Science Framework . ¹ |

The QA rating system is based on the Evaluation Security tool presented in the TASO Monitoring and Evaluation Framework.²

| QA | Comments | Rating (out of 5) |
|-------------|--|-------------------|
| Design | As a quasi-experimental design this project will seek to provide Type 3 evidence on the impact of the Forward Thinking (FT) programme, although the design will not be able to totally account for differences between those who do and don't attend the programme | 4 |
| Sample size | By using multiple years of programme data the sample of approximately 700 FT participants will be matched to a larger pool of students in the HESA data. This is the maximum possible sample and a reasonable sized group for the specified analysis. | 3 |

¹ <https://osf.io/3kjsxw>

² <https://taso.org.uk/evidence/evaluation/>

| | | |
|-----------------|---|------------|
| Outcome measure | The outcome measures as defined map onto the aims of the FP programme. They are 'real-world' outcomes drawn from administrative data and | 5 |
| Attrition | Fuzzy matching will be used to track FT participants in the HESA data. Assuming the FT participant data is sufficiently accurate then attrition in follow-up should be low (assume less than 20%) | 4 |
| Validity | This project is reliant on using existing FT participant data and HESA data to identify comparison groups. Because the data fields do not match-up exactly (e.g. we do not have a perfect 'eligibility' flag in the HESA data) some of the assumptions used to derive this group present inherent limitations to the validity of the comparison. We also do not have data on what other interventions the comparison group might have participated in. These limitations should be made clear in discussion of the results. | 2.5 |
| Overall | | 3.7 |

1. Summary

Background

The Centre for Transforming Access and Student Outcomes in Higher Education (TASO) has funded the University of Birmingham to run an evaluation of their multi-intervention outreach and mentoring programme - Forward Thinking - which has the goal of encouraging more students from widening participation backgrounds to progress to research-intensive universities.

Aims

This project aims to investigate the impact of the Forward Thinking programme at the University of Birmingham.

Intervention

The primary aim of the Forward Thinking programme is to encourage progressions to more research-intensive higher education (HE) providers (those with higher entry requirements). The programme is for 12 to 16 year olds and comprises: a launch event; university experience days; subject taster sessions; mentoring; and a graduation and celebration event.

Design

Higher Education Statistic Agency (HESA) data will be used to track students who have participated in the Forward Thinking programme into HE. Three topics of interest will be investigated:

- Topic 1: Identify other schools in the Birmingham area that had students enter higher education who may have been eligible to participate in the Forward Thinking programme but did not have students attend.
- Topic 2: A deep dive into Forward Thinking students who had entered higher education, investigating their HE journey and outcomes.
- Topic 3: A comparison of Forward Thinking students with a matched group of students (who are assumed not to have participated in the programme) and their HE journey. Where possible, the matched group of students will be similar on available observable characteristics.

Outcome measures

Specifically for Topic 2 (focused on Forward Thinking students who had entered higher education) and Topic 3 (comparing Forward Thinking students who had entered higher education with the matched group of students) - the outcome measures are:

- attendance at a research-intensive HE provider
- the subject they studied (STEM vs non-STEM)
- continuation from first year of study to second year of study

- achieving a first or upper second class honours following completion of a first degree qualification
- time to completion (up to 4/5 years) of first degree study
- progression to postgraduate study following undergraduate study

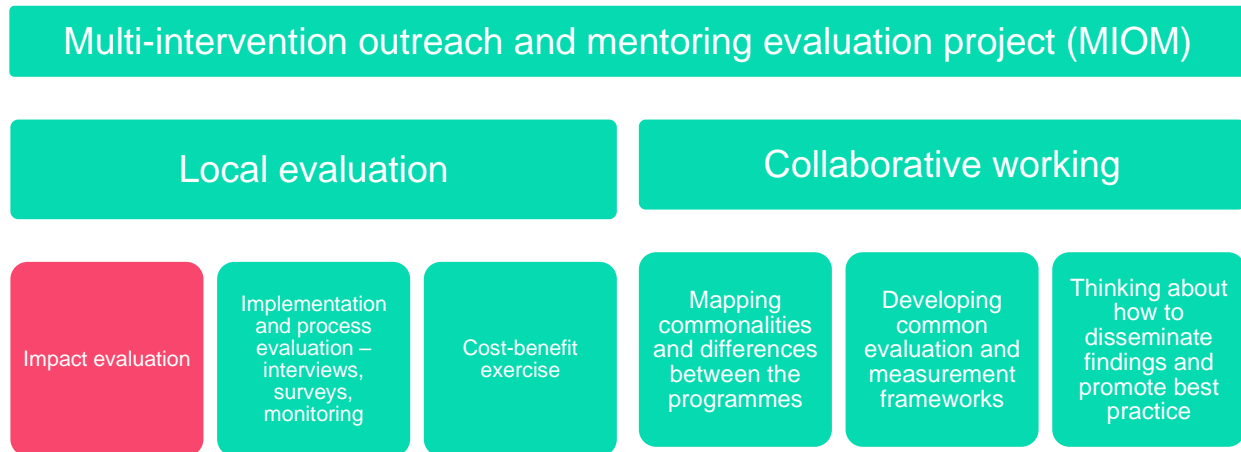
Analyses

Descriptive statistics, tables and charts will be used in analysis for Topic 1 and Topic 2.

For Topic 3, propensity score matching (PSM) using the nearest neighbour method will be employed without replacement to provide matched 1:1 or k:1 comparison groups for Forward Thinking students and non-Forward Thinking students.

2. Background

This evaluation study is part of the TASO-funded project to develop our understanding of multi-intervention outreach and mentoring (MIOM) – that is, programmes which combine multiple outreach strategies into sustained support for learners over a course of months or years. The evaluation study forms one part of a broader evaluation, as shown in the figure below. The parts of the figure which are not highlighted are covered in other planning documents.



The key stakeholders involved in the evaluation are outlined in the table below.

| Organisation | Name | Role and responsibilities |
|--------------------------|--------------------|---|
| University of Birmingham | Elizabeth Chandler | Head of Outreach Overall lead for Birmingham’s involvement in the MIOM project. |
| TASO | Rain Sherlock | Evaluation Manager Lead on the MIOM local impact evaluations |
| Jisc | Emma Jones | Senior consultant- business intelligence Main contact/ project lead- reporting lead/ analysis of data |
| | Vicky Duxbury | Senior Data and Analytics Developer - analysis of data |
| | Nia Comley | Data Intelligence Analyst- leading analysis including linking FT data to HESA data, preparing dataset for analysis and analysis of data |

3. Aims and research questions

The objective of this project is to investigate the impact of participation in the Forward Thinking (FT) programme on a number of outcomes following entry into higher education (HE).

HESA Student Record or Student Alternative data will be used to track students who have participated in the Forward Thinking programme into higher education.³ Three topics of interest will be investigated as follows.

Topic 1: Identify other schools in the Birmingham area that had students enter HE who may have been eligible to participate in the Forward Thinking programme but did not have students attend

Local schools in Birmingham who have had students enter higher education will be identified. Key demographic characteristics about the students from each school will be analysed including entry tariff,⁴ parental education, whether they are from POLAR 1/2 quintile areas, disability status, sex, age on entry, socio-economic classification⁵ and ethnicity. A comparison with the FT cohort of students on these key demographic characteristics will be shown.

Not all criteria for eligibility for the FT programme (defined on page 15) are available via the HESA Student Record, therefore a proxy for the number of eligible students from other local schools will be defined as the sample meeting all the following conditions:

- Achieved a high tariff [relative to the average tariff score of students in the Forward Thinking cohort] AND
- Indicated that their parents did not have a higher education qualification such as a degree, diploma or certificate of higher education AND
- Domiciled from a Polar 1/2 quintile (low progression area) OR indicated they have a known disability)

The total number of eligible students by school will be shown. Providers will not be published by name to ensure anonymity.

Topic 2: A deep dive into Forward Thinking students who had entered higher education, investigating their experience and outcomes

Detailed information about the analysis questions can be found in Section 11, and will include entry tariff distribution of the FT students; analysis of first year FT students such as type of provider attended, continuation from first year of study to second year of study and subject studied; achievement of FT students who had studied a first degree qualification

³ <https://www.hesa.ac.uk/collection/c20051> and <https://www.hesa.ac.uk/collection/c20054>

⁴ Tariff is defined by HESA here: <https://www.hesa.ac.uk/support/definitions/students#tariff>

⁵ Socio-economic classification is defined by HESA in line with the National Statistics Socio-economic classification (NS-SEC), a system based on occupational class - here: <https://www.hesa.ac.uk/support/definitions/students#socio-economic-classification>

and time to completion (start to end of course), and progression to postgraduate study following award of a first degree.

Topic 3: A comparison of Forward Thinking students with a matched group of students (who are assumed not to have participated in the programme) and their higher education experience.

A matched groups design in which the outcome of the treatment group (those students who have participated in the FT programme and have entered HE) are compared to a non-treatment group (those who are assumed not to have participated in the FT programme and have entered HE). The non-treatment group will be selected from HESA Student Record or Student Alternative Record data.

Similar analysis to Topic 2 will be undertaken.

A significance test between the two comparator groups will be undertaken for each outcome measure to see if there are any significant differences in proportions between the two cohorts at a 95% confidence level.

4. Intervention

The intervention being evaluated is the Forward Thinking programme, run by the University of Birmingham. It is a progressive programme of activities for students from Year 8 through to Year 11 who are currently attending a Forward Thinking partner school and comprises:

- Launch event (Year 8)
- Subject taster day (Year 9)
- University experience day (Year 9)
- Mentoring (Year 10)
- Celebration event (Year 11)

A theory of change for this programme is given as [Annex B](#) in Section 14.

5. Design

TASO has commissioned Jisc to undertake the specified analysis using HESA data.

Information about eight cohorts of students who were in Year 8 between 2007/08-2014/15 [n= 701] and had participated in the FT programme will be supplied to Jisc from the University of Birmingham. If these students followed a typical pathway through secondary education and/or further education into HE, it would be expected that the earliest cohort of students would enter HE in 2013/14.

Jisc will attempt to find the FT students in the HESA Student Record 2013/14- 2020/21 through a process using a data linking technique known as fuzzy matching⁶. The HESA Student Record contains information about the academic career of students prior to their enrolment in HE and their achievement at HE. Data linking uses the following information:

- First name
- Last name
- Date of birth (DOB)
- Postcode of domicile

Each student passes through a series of linking 'pots' from 1 to 5, if a student is linked to a pot they will be removed from the next, taking a top down approach. The pot criteria is:

1. First and Last Name/ DOB /Postcode of domicile
2. Soundex name⁷/ DOB/ Same sector postcode of domicile
3. First and Last Name / Postcode of domicile
4. First and Last DOB
5. Same Initial First Name /Last Name/DOB/ Same sector Postcode of domicile

For a student to be linked to a pot they must match every criteria. The strictness of the match reduces from pot 1 to 5. A 10% sample is taken from each pot and a confidence test is undertaken on the number of links that Jisc have complete confidence in and the number that appear ambiguous. The confidence score and number of links from each pot will be shared with TASO and the University of Birmingham. There will be the option to remove pots if the confidence is deemed too low.

FT students may be identified as entering UK HE multiple times in the same or different academic years, for example if they begin and complete a course and enter a subsequent course in a later academic year or if they begin multiple courses in the same or different academic years. Each enrolment is counted separately and is referred to as an instance; an instance is an engagement with a HE provider that aims towards gaining a qualification or credit.

Data linking of the FT students with the HESA Student Record begins with identifying all relevant student instances across HESA Student Records 2013/14-2020/21. If multiple instances are found for a student, Jisc will use a deduplication method to identify the most relevant link, which has a top down approach such that the link which is in the earliest academic year (their first identified engagement with HE) is taken. If there are still multiple links, Jisc will take the link which identifies the best undergraduate qualification (choose first

⁶ Pupils who participated in the FT programme in 2014/15 would be expected to enter HE in 2020/21 at the earliest- these pupils may be less likely to be found in the HESA Student Record or Student Alternative Record as it does not allow for them taking gap years/ repeating years in further education as it does for earlier cohorts of the FT programme.

⁷ Soundex is explained here:

<https://docs.microsoft.com/en-us/sql/t-sql/functions/soundex-transact-sql?view=sql-server-ver15>

degree over other undergraduate) and take a full-time enrolment over a part-time enrolment. This will identify the number of FT students who have entered UK HE and the academic year of entry. For those FT students who have been identified entering into HE, their journey through HE is followed by tracking them through successive years using a personal identifier field (PID)⁸. When the FT students who had entered HE have been identified, the investigation of Topic 2 can be undertaken.

In order to undertake the analysis of Topic 1 and 3, a comparator group of students (non-FT students) from the HESA Student Record 2013/14-2020/21 who were not identified in the FT cohorts of students, and therefore assumed not to have participated in the FT programme, will be extracted. Similarly, their journey through HE is followed by tracking them through successive years using PID.

Specifically for Topic 3, differences in outcome analysis may be due to differences in a background characteristic rather than the study programme that they undertook. A technique known as propensity score matching (PSM)⁹ (see Section 11 for more detail) will be used to create the comparator group of non-FT students. PSM attempts to create similar groups of FT and non-FT students, based on their background characteristics, prior to comparing their outcomes.

Variables within PSM (i.e. the background characteristic variables) are theoretically driven from educational expert knowledge (literature) but are also limited to the availability of the variables that were collected in the HESA Student Record. Possible background characteristics that will be used in the matching algorithm are (listed in priority order):

- Domiciled West Midlands
- Tariff
- Parental education¹⁰
- POLAR 4 quintile
- Disability status

And possibly:

⁸ PID (Personal identifier) is a field used to associate multiple instances of study for a student as we have no way of directly identifying if one instance relates to another in the data. The PID is developed by using fuzzy matching techniques to link all students' instances to a central ID primarily based on underlying fields - First Name, Last Name, Date of Birth, Postcode of domicile and Sex. Student instances do not need to exactly match all the criteria to account for typing errors, change of address and naming differences, thus the PID method is not 100% accurate and should be used with care.

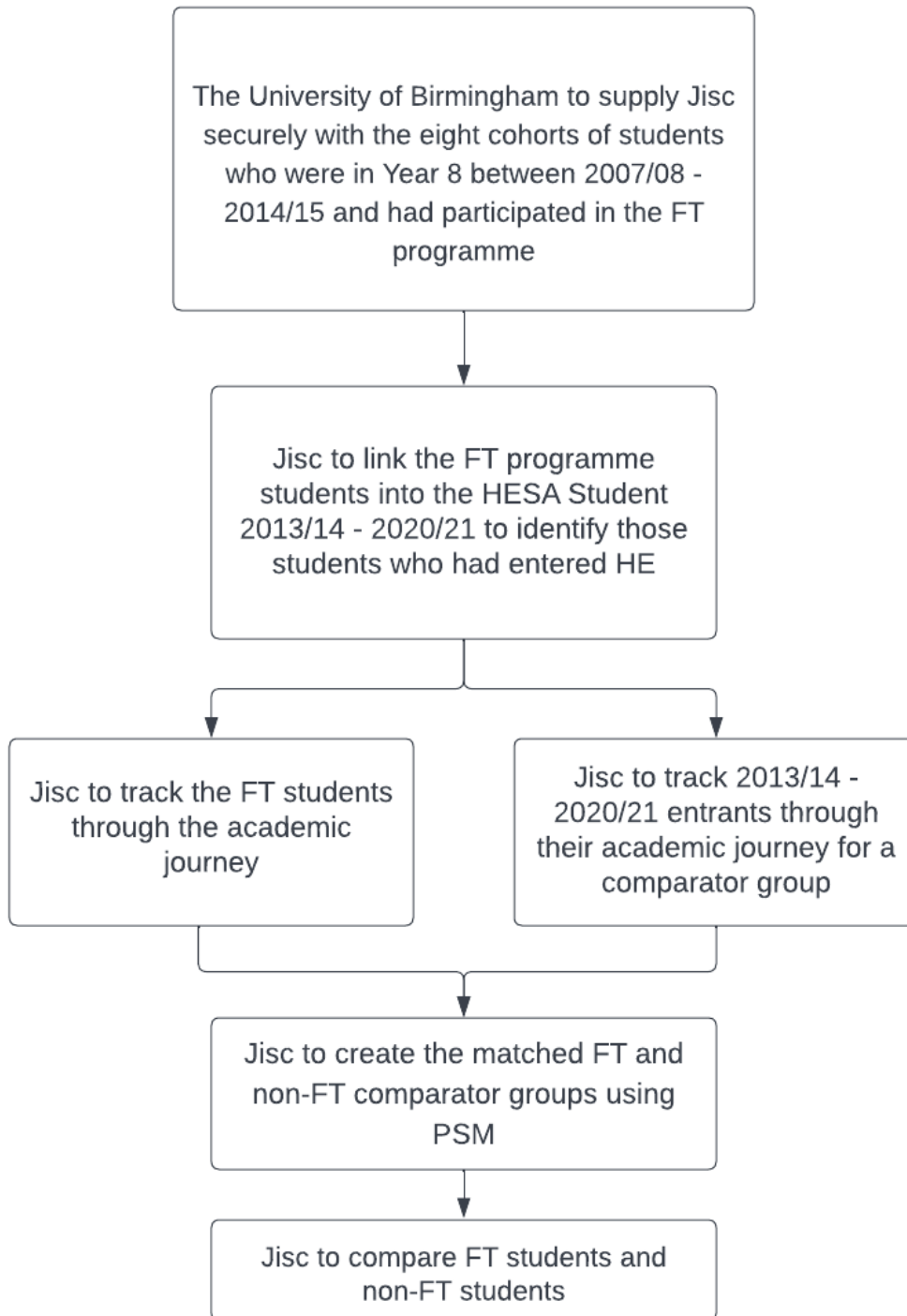
⁹ Rosenbaum, P. R. and Rubin, D. B. 1985. Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score. *The American Statistician*, **39**(1), pp. 33-38 & Rosenbaum, P. R., and Rubin, D. B. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70**(1). pp.41-55.

¹⁰ Parental education is defined here: <https://www.hesa.ac.uk/support/definitions/students#parental-education>

- Age on entry (17 & under/ 18-20/ 21-24/ 25-29/ 30 & over/ Unknown)
- Socio-economic background (classified according to the National Statistics Socio-economic classification (NS-SEC), a system based on occupational class)
- Academic year of entry (allows for unobservable differences between the FT and non-FT students which are common to cohorts such as exam conditions in specific academic years)
- Ethnicity (full)
- State/ private school

Background characteristics will be taken from the HESA Student Record for both FT and non-FT students to ensure that they have been collected consistently using the same methodology.

The propensity score matching will be reprocessed, but amending the domicile matching to England to make a national comparison.



6. Outcome measures

The table lists the analysis of interest which will be investigated via Topic 2 and 3.

| Analysis of interest | Data to be collected | Point of collection |
|---|----------------------|--|
| <p>Primary: Whether the learner attended a research - intensive HE provider (a HEP committed to research as a central part of its mission)</p> | Binary (yes/no) | <p>Via the HESA Student Record and Student Alternative Record</p> <p>(See Annex A for list of research-intensive HE providers)</p> |
| <p>Secondary: Whether the learner studied a STEM subject</p> | Binary (yes/no) | <p>Via the HESA Student Record and Student Alternative Record</p> <p>(See footnote 14 for STEM classification)</p> |
| <p>Secondary: Whether the learner continued from 1st year to 2nd year of study</p> | Binary (yes/no) | Via the HESA Student Record and Student Alternative Record |
| <p>Secondary: For those learners who completed a first degree, whether the learner achieved a first or upper second class honours first degree</p> | Binary (yes/no) | Via the HESA Student Record and Student Alternative Record |
| <p>Secondary: Whether the learner completed their qualification in up to 5 years</p> | Binary (yes/no) | Via the HESA Student Record and Student Alternative Record |

| | | |
|---|-----------------|--|
| Secondary: Whether the learner progressed to postgraduate study | Binary (yes/no) | Via the HESA Student Record and Student Alternative Record |
|---|-----------------|--|

7. Sample selection

The study will include eight cohorts of students who started Year 8 at school (aged 11 and 12 years old) between 2007/2008 and 2014/2015 and who completed the FT programme. Of the learners, 313 (44.7%) were male, 284 (59.5%) were BAME, the majority (97%) did not have a parent or carer who had previously attended HE, 233 (41.5%) were eligible for free school meals. 71.0% of learners were from POLAR quintiles 1 and 2 (table 1)

Table 1: Demographic characteristics of FT learners

| Cohort (start year) | Sample size | Male | BAME | No parental HE | FSM | POLAR 1 or 2 |
|---------------------|-------------|-------------|-------------|----------------|-------------|--------------|
| 1 (2007/08) | 45 | 23 (51.1%) | 17 (37.8%) | 41 (91.1%) | Unknown | Unknown |
| 2 (2008/09) | 71 | 32 (45.1%) | 27 (38.0%) | 69 (97.2%) | 21 (29.6%) | 54 (76.1%) |
| 3 (2009/10) | 48 | 20 (41.7%) | 41 (85.4%) | 47 (97.9%) | Unknown | 35 (72.9%) |
| 4 (2010/11) | 47 | 18 (38.3%) | 36 (76.6%) | 45 (95.7%) | Unknown | 32 (65.1%) |
| 5 (2011/12) | 85 | 37 (43.5%) | 43 (50.6%) | 81 (95.3%) | 26 (30.6%) | 65 (76.5%) |
| 6 (2012/13) | 116 | 53 (45.7%) | 64 (55.2%) | 112 (96.6%) | 44 (37.9%) | 76 (65.5%) |
| 7 (2013/14) | 159 | 67 (42.1%) | 93 (58.5%) | 157 (98.7%) | 75 (47.2%) | 78 (49.1%) |
| 8 (2014/15) | 130 | 63 (48.5%) | 96 (73.8%) | 128 (98.5%) | 67 (51.5%) | 126 (96.9%) |
| Total | 701 | 313 (44.7%) | 417 (59.5%) | 680 (97%) | 233 (41.5%) | 466 (71.0%) |

In order to take part in the FT programme learners had to have met one of the following criteria: 1) Have the academic potential to achieve good GCSE grades and go to a research intensive University and; 2) Have parents/guardians who have not completed a HE qualification in the UK or abroad.

In addition, students should meet at least one of the following criteria:

- Live in a low progression area
- Eligible for pupil premium funding or free school meals
- Has a recognised disability
- Has experienced significant extenuating circumstances that has had (or is having) a detrimental impact on their studies
- Be a young carer

8. Data collection

All FT learner data has already been collected. HESA data is accessible via Jisc. The variables from each dataset required are listed in Table 2 below.

Data storage and security

The FT data will be supplied to Jisc via a secure upload link with access to the data limited to members of the team who require access for linking. This is stored in a password protected secure area. Privacy notices and consent forms have been assessed by the Jisc Data Protection Officer. A data sharing agreement has been put in place between Jisc and The University of Birmingham. Once the project has concluded this data will be deleted.

The linked FT data and HESA data does not include the student name, date of birth or postcode information - a unique identifier is used to identify individualised students within the dataset.

All members of the Jisc team have undergone annual Data Protection and Information Security training.

The dataset will not be supplied to TASO nor The University of Birmingham and will be solely used for the purposes of this project. All data used within reports will be rounded using the [HESA Rounding Strategy](#).

Table 2- Data items

| Data item | Timeframe | Collector |
|---|---|--|
| <p>Learner information</p> <ul style="list-style-type: none"> • Cohort • First name • Last name • Date of birth (DOB) • Postcode of domicile | <p>2007/08-2014/15</p> | <p>The Forward Thinking Team</p> |
| <p>Learner and Comparator Group Data¹¹</p> <ul style="list-style-type: none"> • Individualised student data • Academic year <p>Information prior to entering HE, including background characteristics of HE entrants:</p> <ul style="list-style-type: none"> • Sex • Age on entry (17 & under/ 18-20/ 21-24/ 25-29/ 30 & over/ Unknown) • Low participation neighbourhood marker (POLAR4 Quintiles) • Ethnicity (White/ Black - Caribbean/ Black - African/ Other Black background/ Asian - Indian/ Asian - Pakistani/Asian - Bangladeshi/ Chinese/ Other Asian background/ Mixed White and Black / Mixed White and Asian/ Other mixed background/ Other/ [Unknown/Not applicable]) • Care leaver (Care leaver/ Not a care leaver) • Disability marker (Known disability/ No known disability) • Tariff • Highest qualification on entry • Parental education • Socio-economic Classification • State school marker • Domicile (UK/ EU/ Non-EU/ Unknown) • Domicile (Local authority district) • Domicile (UK county/ Non-UK country/ Unknown) | <p>2013/14-2020/21¹² for HESA Student Record and 2014/15–2020/21 for HESA Student Alternative Record</p> | <p>HESA: https://www.hesa.ac.uk</p> |

¹¹ Hyperlinks to definitions included where available

¹² HESA have collected data from the sector since the 1994/95 academic year.

| | | |
|--|--|--|
| <ul style="list-style-type: none"> • Last provider attended (postcode/ local authority/ name) <p>Other data to be used for outcomes analysis</p> <ul style="list-style-type: none"> • HE provider • Mode of study (Full-time/ Part-time) • First year marker • Term time accommodation [not available for HESA Student Alternative Record] • Expected length of programme • Subject of study 1 (CAH1) • Subject of study 2 (CAH1) • Subject of study 3 (CAH1) • Subject of study 4 (CAH1) • Subject of study 5 (CAH1) • Subject of study balance indicator (CAH1) • Subject of study 1 percentage (CAH1) • Subject of study 2 percentage (CAH1) • Subject of study 3 percentage (CAH1) • Subject of study 4 percentage (CAH1) • Subject of study 5 percentage (CAH1) • STEM subject marker • Level of study (Postgraduate taught/ Other postgraduate/ First degree/ HND/ Foundation degree/ Other undergraduate) • Continuation status <p>If progress from 1st to 2nd year:</p> <ul style="list-style-type: none"> • Progression (Yes/no) • Still active in HE (Yes/no) • Year of programme last found active • Change in HE provider (Yes/no) <p>If completed qualification:</p> <ul style="list-style-type: none"> • Academic year of qualification • Class of first degree • Level of qualification (Postgraduate taught/ Other postgraduate/ First degree/ HND/ Foundation degree/ Other undergraduate) • Time to completion (End date- start date) <p>If on completion, progress to postgraduate level study:</p> <ul style="list-style-type: none"> • Level of study (Postgraduate taught/ Postgraduate research) • HE provider • Gap between courses (years) • Subject of study 1 (CAH1) | | |
|--|--|--|

| | | |
|---|--|--|
| <p>Last found in HESA Record information:</p> <ul style="list-style-type: none"> ● Still active in HE (Yes/no) ● Year of programme last found active ● End date (if returned) ● Still active on original course (based on course ID) | | |
|---|--|--|

9. Procedure

A project timeline is provided in the table below:

| Timeframe | Action |
|---|--|
| October 2022 | <ul style="list-style-type: none"> ● Data sharing agreement in place between Jisc and The University of Birmingham ● Delivery of FT data from The University of Birmingham to Jisc |
| November - December 2022 | <ul style="list-style-type: none"> ● Research protocol finalised |
| December 2022 | <ul style="list-style-type: none"> ● Analysis for Topic 1 and 2 |
| January - February 2023 (timeline dependent on when redrafted research protocol approved) | <ul style="list-style-type: none"> ● Analysis for Topic 3 ● Report writing |
| February 2023 | <ul style="list-style-type: none"> ● Delivery of draft report to TASO |
| March 2023 | <ul style="list-style-type: none"> ● Edits and report sign-off |

11. Analytical strategy

Topic 1

Summary tables will show number of pupils from other schools in the Birmingham area by key demographic characteristics including, but not limited to, sex, ethnicity, known disability, parental education etc.

This information will also be shown for the FT cohort of students.

The number of eligible students from other local schools will be defined as (a proxy of eligibility criteria):

- Achieved a high tariff [relative to the average tariff score of students in the Forward Thinking cohort]
- Indicated that their parents did not have a higher education qualification such as a degree, diploma or certificate of higher education [Use PARED field]

- Domiciled from a Polar 1/2 quintile (low progression area) OR indicated they have a known disability [Use DISABLE field]

The total number of eligible students by school will be shown. Information about the school, such as school type, school capacity and percentage of learners receiving Free School Meals (FSM) will be accessed from [government schools data](#). Schools who are not eligible to participate in the FT programme such as sixth form colleges or selective schools will be excluded.

Topic 2

The following analysis will be undertaken and be displayed via charts and summary statistics; where relevant, comparisons may be made to the complete HESA Student population:

- Percentage of Birmingham FT students found in the HESA Student Record and Student Alternative Record (progression to HE)
- Overall numbers over time
- Due to small numbers further analysis will be conducted combining all years of study
 - What is the entry tariff distribution of the FT students?
 - What proportion of first year FT students attended a research-intensive HE provider?¹³ Analysis will be undertaken by level of study and mode of study.
 - What proportion of first year FT students studied a STEM¹⁴ subject? Analysis will be undertaken by attendance at a research-intensive HE provider.
 - What proportion of FT students continued from first year of study to second year of study? Analysis will be undertaken by attendance at a research-intensive HE provider and STEM subject study.
 - What proportion of FT students achieved a first or upper second class honours following completion of a first degree qualification? Analysis will be undertaken by attendance at a research-intensive HE provider and STEM subject study.

¹³ A research-intensive provider is defined as a Russell Group HE provider:

<https://russellgroup.ac.uk/about/our-universities/> and former members of the 1994 Group. A full list of HE providers can be found in Annex A in section 14 of this protocol.

¹⁴ The STEM grouping includes all Common Aggregation Hierarchy (CAH) level 1 codes CAH01 through to CAH13 and CAH26 with the exception of CAH26-01-03 (Human geography). CAH26 (Geographical and environmental studies) has been disaggregated so that CAH26-01-03 (Human geography) is presented in the non-science grouping labelled as 'Geographical and environmental studies (social sciences)'. All other CAH level 3 codes within CAH26 are presented in the STEM grouping labelled as 'Geographical and environmental studies (natural sciences)'. This grouping of STEM subjects has been created by HESA. <https://www.hesa.ac.uk/support/documentation/hecos/cah#download-cah>

- The proportion of FT students by time to completion (up to 4/5 years) of those who studied a first degree study. Analysis will be undertaken by attendance at a research-intensive HE provider, STEM subject study and class of degree marker
- What proportion of FT students who had completed a first degree then progressed to postgraduate study? Analysis will be undertaken by attendance at a research-intensive HE provider, STEM subject study and class of degree marker

Topic 3

In order to create a matched cohort of students who did (assumed) not to participate in the FT programme PSM using the nearest neighbour method will be employed without replacement to provide matched 1: 1 or k:1 comparison groups for FT students and non-FT students (strategy described in Section 5). It is expected that exact matches will be established due to the large amount of data collected via the HESA Student Record and Student Alternative Record , though different techniques will be considered including exploring matching distances such as Mahalanobis distance. Exploratory analysis will be undertaken, measuring precision against robustness to determine the appropriate value of k used if k:1 matching is to be undertaken. A variety of post-match balance tests will be undertaken to assess the effectiveness of the matching process.

Similar charts/ analysis as described in Topic 2 will be produced. A significance test between the two comparator groups will be undertaken on each outcome to see if there are any significant differences in proportions between the two cohorts at a 95% confidence level.

12. Ethical considerations

This study has been reviewed by and received a favourable opinion from the University of Birmingham's Humanities and Social Sciences Ethics Committee at the University of Birmingham (reference: ERN_21-0539).

13. Risks

| Part of evaluation | Risk | Mitigation strategy | Risk owner |
|-------------------------------|---|--|-------------|
| Results/findings | That any results from the analysis are taken out of context, given the limitations of the design | The analytical report and any final deliverables will clearly outline the limitations of the study and evaluation design. | TASO & Jisc |
| Forward Thinking data quality | Poor data capture such as spelling errors/ partial postcodes leading to poor quality matches in the HESA data | <p>Data linking using a technique known as fuzzy matching will attempt to find the FT students in the HESA Student Records. As detailed on page 9, data linking uses the following information:</p> <ul style="list-style-type: none"> ● First name ● Last name ● Date of birth (DOB) ● Postcode of domicile <p>Each student passes through a series of linking 'pots' from 1 to 5, if a student is linked to a pot they will be removed from the next, taking a top down approach. The pot criteria is:</p> <ol style="list-style-type: none"> 1. First and Last Name/ DOB /Postcode of domicile 2. Soundex name / DOB/ Same sector postcode of domicile 3. First and Last Name / Postcode of domicile 4. First and Last DOB 5. Same Initial First Name /Last Name/DOB/ Same sector Postcode of domicile <p>For a student to be linked to a pot they must match every criteria. The strictness of the match reduces from pot 1 to 5. A 10% sample is taken from each</p> | TASO & Jisc |

| | | | |
|---|-----------------------------|--|-------------|
| | | pot and a confidence test is undertaken on the number of links that Jisc have complete confidence in and the number that appear ambiguous. Student links from pots 1-5 with a confidence score of more than 80% will be used in the analysis. | |
| Forward Thinking student linking to HESA data | Low or no matches are found | The University of Birmingham has followed up with some Forward Thinking students who have entered HE. We will be able to check that the matching is, at least to some degree, successful. Appropriate analysis will be undertaken given the number of matches found. | TASO & Jisc |

14. Annexes

Annex A: HE providers who are defined as ‘research-intensive’. This list aligns with the definition used by [Realising Opportunities](#).

1994 Group

0127 Birkbeck College

0117 The University of East Anglia

0118 The University of Essex

0131 Goldsmiths College

0133 Institute of Education

0123 The University of Lancaster

0125 The University of Leicester

0152 Loughborough University

0141 Royal Holloway and Bedford New College

0146 SOAS University of London

0162 The University of Sussex

Russell Group

0110 The University of Birmingham

0112 The University of Bristol

0114 The University of Cambridge

0179 Cardiff University

0116 University of Durham

0167 The University of Edinburgh

0119 The University of Exeter

0168 The University of Glasgow

0132 Imperial College of Science, Technology and Medicine

0134 King's College London

0124 The University of Leeds

0126 The University of Liverpool

0137 London School of Economics and Political Science

0204 The University of Manchester

0154 Newcastle University

0155 University of Nottingham

0156 The University of Oxford

0139 Queen Mary University of London

0184 Queen's University Belfast

0159 The University of Sheffield

0160 The University of Southampton

0149 University College London

0163 The University of Warwick

0164 The University of York

Annex B: The Theory of Change displayed has been designed at the programme level, to illustrate the impact of Forward Thinking. Within the delivery of Forward Thinking, there is a focus on STEM subjects - not captured by this ToC.

