Breaking Barriers
Meaningful employment for refugees

Reaching meaningful employment

Understanding the impact of Breaking Barriers' one-to-one model of employment support for refugees

Contents

Introduction

About Breaking Barriers

Breaking Barriers exists so every refugee can access meaningful employment and build a new life. We welcome refugees into meaningful employment with advice, experience, and education. We believe in the power of responsible business to change society for the better through our innovative partnerships.

As a specialist refugee employment charity, our programmes are tailored to our clients to support them to tackle the multiple barriers to employment and integration they're facing, including gaps on CVs caused by the asylum process, lack of knowledge of UK hiring and workplace culture, and employer misperceptions about hiring refugees.

We deliver one-to-one employment advice and guidance, and education courses to tackle the underlying barriers to employment – such as insufficient English language and digital skills. Finally, we work directly with over 50 businesses to deliver skills workshops, paid work placements, and job opportunities for refugees.

Using our impact and data-led approach, we seek to influence wider change within the refugee sector and beyond by sharing our research and insights. Ultimately, our aim is to address the complex challenge of refugee unemployment by providing holistic front-line support, while working towards long-term systemic change for refugees in the UK.

Refugees in the UK and barriers to employment

As of November 2022 there were 231,597 refugees, 127,421 pending asylum cases and 5,483 stateless persons in the UK (1). The war in Ukraine has driven a large increase from the previous year which was roughly 130,000 refugees.

Just over three quarters (76%) of the asylum decisions in 2022 were grants (of refugee status, humanitarian protection, or alternative forms of leave), which is a substantially higher grant rate than in pre-pandemic years and the highest yearly grant rate since 82% in 1990.

Of the top 10 nationalities applying for asylum, half have a grant rate above 80% (Afghanistan 98%, Iran 80%, Syria 99%, Eritrea 98%, and Sudan 84%) (2).

Once refugee status is granted, refugees living in the UK face many challenging barriers to integrate into the UK, one of the main ones being access to employment. These barriers include language barriers, non-recognition of existing qualifications, cultural differences, and racism and discrimination.

1 https://www.unhcr.org/uk/asylum-in-the-uk.html, April 2023

2 https://www.gov.uk/government/statistics/immigration-system-statistics-year-ending-december-2022/summary-of-latest-statistics

84% of refugees reported that they did not have sufficient English language ability to get a job (3). Support for these issues is very limited – in some places, waiting lists for English classes are two years long, and the majority of those in classes say that the classes they are doing are not sufficient to learn the language.

As a result, refugees in the UK are 4 times more likely to be unemployed than people born here, and on average earn about half the amount per week that UK nationals do. This is despite high levels of qualifications and skills (38% of refugees from Syria living in the UK have a university degree (4), for example).

3 https://www.deloitte.com/global/en/issues/work/talent-displaced.html

4 https://www.deloitte.com/global/en/issues/work/talent-displaced.html

Making our research accessible

The data analysis in this report describes a statistical approach for evaluating Breaking Barriers' work. However, we want to make this report accessible for as wide a readership as possible, while also providing sufficient levels of technical detail for those that want it.

Consequently, this report is not written in an academic style. We hope that any readers without a statistical background who are interested in Breaking Barriers will be able to read this report and take away the key messages.

You may find our previous report Effective employment support for refugees: Breaking Barriers approach a useful read as it offers an overview of our programmes, whereas this report deep dives into one aspect of our model; one-to-one support.

Throughout the report, we have also included information boxes that introduce some of the concepts used in the analysis, as well as contextual information. More detail on the technical aspects of the analysis can be found in the footnotes or in the appendix.

If you have any further questions about the report, you can email Toby Gill at t.gill@breaking-barriers.co.uk.

Executive summary

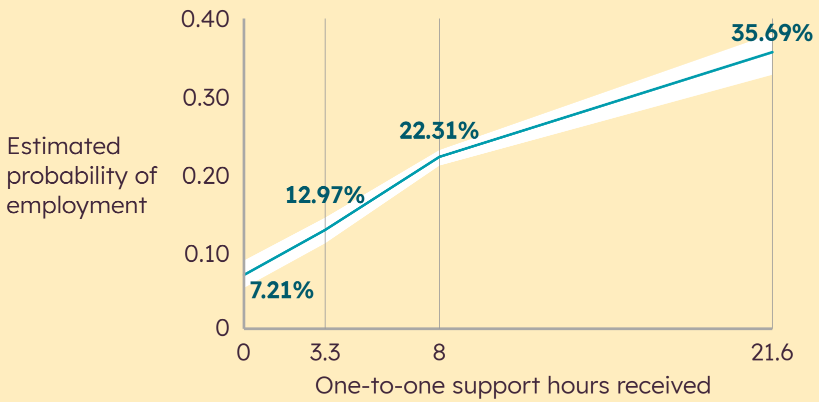
Since launching in 2015, Breaking Barriers has published annually how many of our clients have successfully entered employment. However, we could not precisely estimate how much more likely a client was to get a job because of Breaking Barriers' support. This report is the first attempt to explore this question.

We used statistical modelling to explore how Breaking Barriers' one-to-one support sessions impact the probability of refugees entering employment.

The model estimated that most of the benefit resulting from Breaking Barriers' support is delivered within the first 8 hours of support, with impact per hour reducing after this point.

Other parameters included in the model yielded valuable insights about various factors that influence refugee success in finding employment. For example, we were able to estimate how far a client's age, English language level, previous job-seeking experience, and security of housing impact their probability of finding a job, when controlling for all other factors.

One-to-one support hours and employability



Our model estimated that:

8 hours of one-to-one support sessions **more than tripled a client's chances** of finding employment (from 7% probability to 22%).

80% of all clients' employment outcomes **would not have occurred without the support they received** from Breaking Barriers.

Meet an adviser

I'm Ellie, and I work directly with refugee and asylum-seeking clients, supporting them to work towards their employment and education-related goals. I also lead on supporting our clients who are underemployed.

I have a Masters in Migration, Mobility, and Development, with research focusing on gendered experiences of humanitarianism and borders and have a number of years of experience working in the refugee sector both in the UK and Greece.

As an employment adviser, we spend time supporting clients to work toward their long-term and short-term goals, focusing on manageable actions. This can range from reviewing CVs, cover letters, and job applications, to building confidence and other skills including public speaking. Our clients at Breaking Barriers are all so unique – no one day as an adviser is the same!

Working with clients to help them access the information and support needed to move closer to and achieve their goals is great – I've met so many interesting people, with such amazing experience and skills. I've definitely learned a lot about all sorts of niche topics!

My favourite part of the job is hearing someone say they are proud of themselves. Watching the confidence of someone grow when they're given the space to be listened to and heard is particularly special.

Ellie – Senior employment and integration adviser

How we deliver support

This diagram provides an overview of how refugee clients that need our support access our programme, and how the one-to-one aspect of our support is then realised through the relationship with an adviser.

Referral

Referrals come from other charities, job centres and NHS services

- Client refers themself
- Referral partner refers client

Client joins waiting list →

Enrolment with Adviser

- Discuss professional and educational experience
- Identify support needs outside of employment
- Set short and long term goals and agree support plan

Client joins Adviser's caseload →

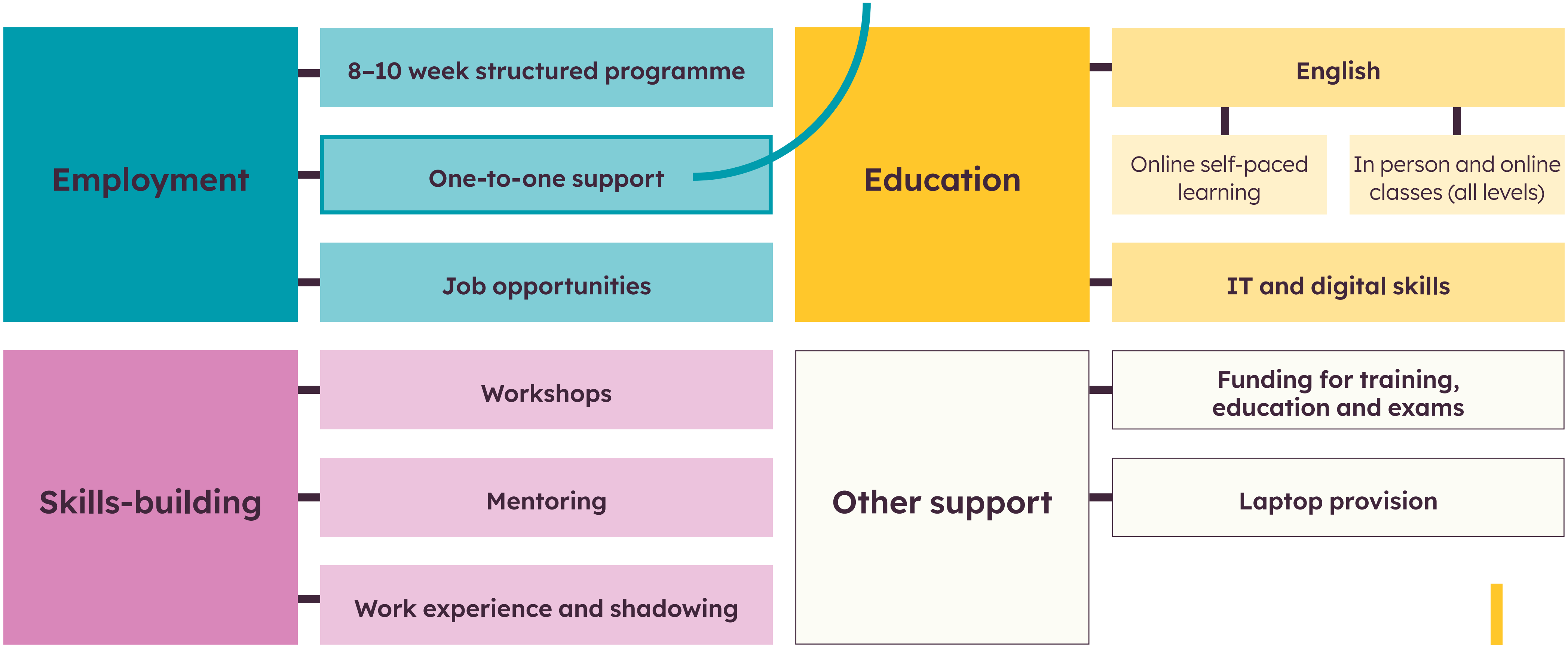
Client's support journey starts

- Adviser provides support and identifies relevant opportunities within Breaking Barriers (for example education courses, workshops, job opportunities)
- Adviser signposts or refers to specialist organisations who can support with needs outside of employment
- Adviser monitors progress, reviews goals / support plan with the client until the client achieves outcome and / or exits programmes

Our programmes

This diagram provides an overview of the Breaking Barriers employment and education programmes and shows how one-to one support is an integral aspect of that, but not all that we offer.

This research is a deep dive into the one-to-one aspect of our employment model’



Our one-to-one model

At the core of Breaking Barriers' support for refugees is one-to-one employment support sessions, which we refer to internally as Information, Advice and Guidance sessions (IAGs).

These are one or two-hour appointments that take place to achieve a specific task that will help a client progress towards a goal. This could be working on a CV, practicing digital skills, making a college or university application, applying for funding, researching a specific sector, or preparing for a job interview. IAGs can take place remotely or in person, depending on the client's preference.

IAG sessions are typically between a refugee client and our expert employment and integration advisors or with our advice and guidance volunteers. Every client is paired with their own dedicated advisor who helps them to clarify their goals, understand the UK labour market, identify job opportunities, and put together applications.

Clients

At Breaking Barriers we use the term 'clients' to refer to the refugees we support with our services. While eligibility for our programmes covers those with a variety of different visas and immigration statuses, broadly speaking all our clients are refugees with the right to work in the UK who would like support in finding the right job for them. We use the term 'client' rather than 'beneficiary' or 'refugee' because it encapsulates our welcoming approach and the professional service offer.

Outcomes

At Breaking Barriers, we monitor the effectiveness of our work by tracking outcomes. An outcome represents a client taking tangible steps toward achieving their career goals. A client achieves an outcome whenever they enter employment (be that a first job, a promotion or a career change), start an academic qualification or begin a training course. For this report, we are only looking at employment outcomes.

Objectives of this research

As mentioned in the previous section, at the core of Breaking Barriers' support for refugees are one-to-one support sessions with expert Employment and Integration Advisors and volunteers.

Although we provide other forms of support (such as education classes covering English language and technical skills, and sector-specific and skills-based workshops with our corporate partners), most of the clients' time with us is via one-to-one support. Understanding the impact of one-to-one sessions is therefore crucial to understanding the efficacy of our model.

We measure the proportion of our refugee clients that enter employment every year. However, to understand the true impact of our support, we need to know how many of these clients would have entered employment without us and how many would not. This proportion is known as our 'attribution rate', and it is fundamental for impact measurement.

However, it is impossible to firmly establish an attribution rate without running a Randomised Controlled Trial (RCT). This is where Breaking Barriers, and many other charities, run into problems – to gather comprehensive data we would need to withhold our services from a control group of refugees, which would be logistically challenging and unethical.

Randomised Controlled Trials and Attribution Rates

Randomised Controlled Trials (RCTs) are a rigorous experiment that demonstrates the impact of a particular intervention. They are designed for testing whether new drugs or medical procedures are effective for patients (but can also be used to test whether charitable programmes are achieving their goals).

In an RCT, patients are sorted randomly into two groups. The 'test' group is given the treatment, while the 'control' group is not (and may be given a placebo instead). In both groups, a number of patients' symptoms will improve. Only by comparing the test group with the control group can the experimenter know how much of the improvement in the test group was caused by the new treatment, rather than by random chance. In evaluations of interventions by charities/non-profits, this is typically referred to as the 'attribution rate' – the amount of improvement that can be attributed directly to the treatment (so, for example, if the attribution rate is only 20%, that means 80% of the patients who recovered would have done so even without the treatment).

For Breaking Barriers, our attribution rate is the percentage of clients, out of those who got jobs, who wouldn't have gotten those jobs without us. It is a vitally important measure of our programme's impact. However, without a control group, knowing an attribution rate for certain is impossible.

This research is an attempt to circumvent this problem. While causality cannot be firmly established without an RCT, we can feed our data into a statistical model. This shows us correlations in our data and estimates the relationship between the amount of support we provide with a probability of our clients finding employment.

Our goal was to fit a statistical model to the Breaking Barriers client dataset and train it to predict the probability of employment outcomes, to:

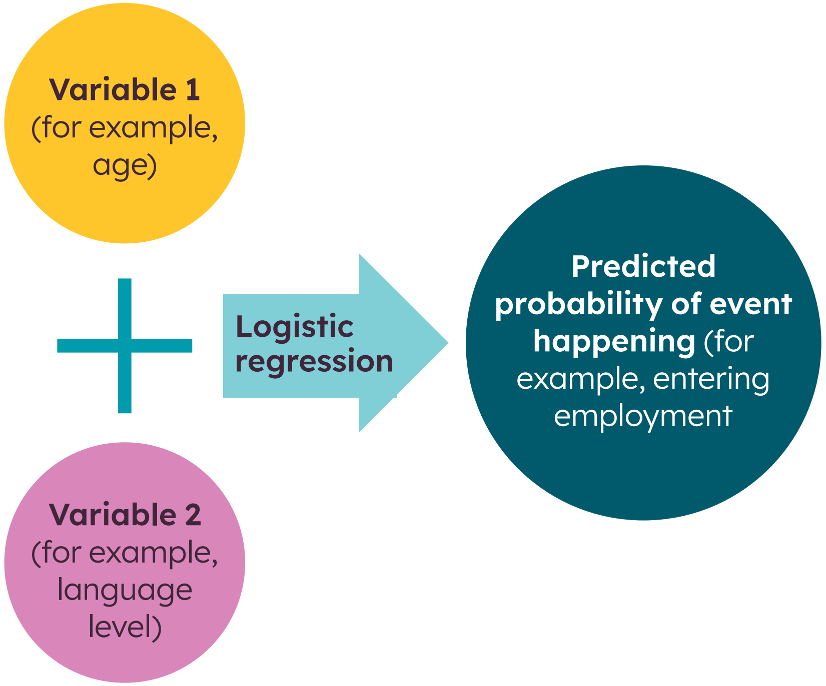
1. Estimate the relationship between the number of one-to-one support sessions a client receives and the probability of them entering employment, and to use this relationship to estimate our 'attribution rate'.

2. Gain an understanding of the role played by other factors (entered as control variables in our model) in helping clients get a job.

Methodology

The analysis deployed is a technique called logistic regression – a classification algorithm that is frequently used in machine learning (5). A logistic regression model takes a selection of variables as inputs and returns an estimated probability of an event happening as its output (in this case, the probability of a client finding employment).

We trained a logistic regression model on our client dataset, which included various demographic traits (such as age, English language proficiency and level of education) and the number of hours that each client had spent receiving one-to-one support. The dataset was limited to contain only clients who were unemployed at the point of enrollment (6). The model was set to predict the probability of clients entering employment during their time on the programme.

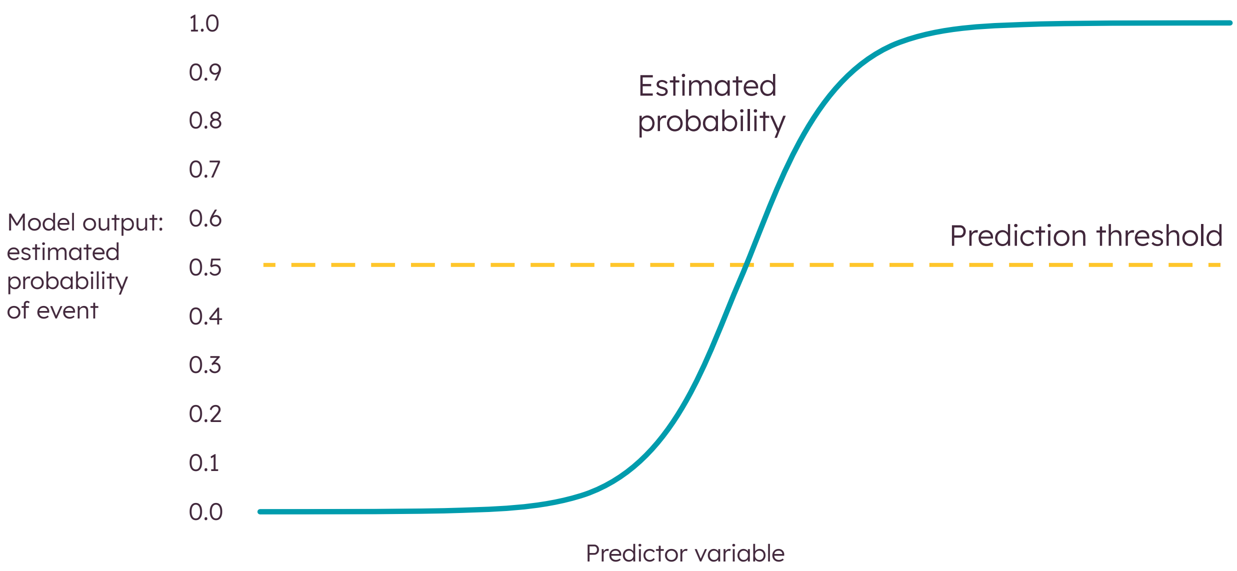


5 Machine learning models come in two types – supervised and unsupervised. A supervised model is taught by example from existing training data, whereas an unsupervised model is given no example answers in the training data and must draw its own inferences. An example of an unsupervised model is a 'clustering' algorithm that sorts data points into groups (which is unsupervised because these groups, or 'clusters', were not present in the data the model was trained on). Supervised models, meanwhile, look at the 'answers' in existing cases and predict what the values should be in new cases. The values being predicted can either be numerical (predicted by 'regression' models), or categories (predicted by 'classification' models). For this project, we wanted to train a model on existing data to predict whether a client would be in the 'gets a job' category or the 'doesn't get a job' category, so we needed a classification model. Logistic regression is one of the most widely used classification models available.

6 Breaking Barriers also supports clients who are already employed but who want to progress their career, however, for the sake of simplicity, these clients were not included in this study. Clients that enrolled before financial year 2019 were excluded as data for several key variables included in the model was not gathered before this date.

The key to making this approach successful was that Breaking Barriers' intervention can come in different 'dosages', i.e., clients can receive varying numbers of support hours. This variety enabled us to explore how, all else being equal, a client with more hours might be more likely to enter employment than a client with less. Once this relationship had been established, we could use the model to hypothesise about what might have happened if all the clients had received zero hours of support (i.e., if Breaking Barriers had not supported that client at all) enabling us to estimate an attribution rate (i.e., the proportion of the clients who entered employment that would not have done so without Breaking Barriers' help).

Example logistic regression model



A logistic regression model estimates the relationship between the predictor variable(s) (hours of support received) and the estimated probability of an event happening (getting a job). As the model's output is an estimated probability, the value is always a number between 0 and 1 (hence the 'S' shape of the curve above, as the probability must flatten before it crosses either 0 or 1). The model is then used to generate an estimated probability for each case in the dataset. A prediction threshold is set (this is done by trialling a variety of different threshlds and seeing which generates the most accurate predictions). If a case has an estimated probability above that threshold, it is predicted that the event will happen (e.g. the client will enter employment). If the estimated probability is below that threshold then it is predicted that the event will not happen.

Meet an adviser

I started working for Breaking Barriers in January 2022, after 3 years working in Scotland for an organisation offering generic support to refugees and asylum seekers nationwide.

During this period, I was overwhelmed with requests from community leaders looking for information and advice for refugees who were struggling to enter the job market. After some research I quickly realised that the support provided towards this specific need was almost non-existent. It was through this research that I found Breaking Barriers and became very interested in their work.

More than one year on, working as an Employment and Integration Adviser has been one of the most fulfilling experiences in my career so far. I get to meet people from different countries and walks of life, which is enriching and inspiring. After discussing my clients' goals and defining an action plan together, I get to see them progressing more and more in their journey toward independence. It's not always easy, and there are many bumps in the road, but it's great to see the clients' confidence increase.

My day-to-day activities revolve around contacting clients for regular check-ins and delivering support sessions. I also organise and facilitate sessions delivered by volunteers, so that as many clients as possible can receive support toward their goals each week. The remaining time is then spent on the following tasks: referring clients to organisations providing different types of support (mental health, housing, immigration, etc.); attending external workshops to gather new information on the sector as well as useful services or initiatives and sharing relevant information with the team and clients; and supporting and facilitating the delivery of internal workshops, career insight events and mentoring opportunities.

Marta – Employment and integration adviser

Limitations

It is important to be mindful of the various limitations and drawbacks of this type of approach, particularly when interpreting the findings of the model. The most important limitations are:

1. This kind of regression analysis cannot firmly establish causation (only a randomised controlled trial can do this). This kind of analysis can only uncover correlations. While correlations can give an impression of underlying trends in the dataset, they cannot establish what caused the trend (for example, it might be that clients completed more support hours because they were more employable, rather than the other way around). For this reason, we can only give an 'estimated' impact on probability of employment, rather than a definitive figure.

2. The model only estimates a client's probability of entering employment. It does not consider clients' other achievements which we also consider as 'outcomes' (such as entering education or training).

3. The model only looked at clients who were unemployed when they enrolled with Breaking Barriers. This is not reflective of all the clients that we support. Some clients are 'underemployed', in that they already have a job but want to move to a role that makes better use of their skills.

4. A considerable amount of our work involves working directly with corporate partners to raise awareness of challenges facing refugees, encouraging firms to hire more refugees and ultimately create job roles and placement opportunities for clients to step into. Unfortunately, this work proved impossible to incorporate into the model, so its impact is overlooked here.

5. Similarly, education classes and workshops proved too challenging to incorporate into the model (partly due to the low number of hours clients complete in workshops relative to regular support sessions, and partly due to the longer timescales involved in teaching clients a new language).

6. Breaking Barriers is often not the only organisation supporting our clients – they receive support from other charities and groups relating to other areas of their lives, such as housing, the asylum process, transport, and health.

7. The model only considers the probability that a client will enter employment. It does not consider the nature of that employment, in terms of seniority, desirability of working conditions or salaries, all of which are vital to how those roles will impact a client's wellbeing.

8. The model is designed to estimate the probability of a client achieving employment during the time they were on a Breaking Barriers programme (i.e., the period we have data for them). It does not predict the probability of them achieving employment ever. It is probable that the clients that the model predicted would not have entered employment without Breaking Barriers' support could have found a job eventually.

Evaluating the model

Searching for employment is an inherently unpredictable process. There are many aspects of employability that are intangible and difficult to capture numerically, and chance events (such as whether there is chemistry between an interviewer and interviewee) can play decisive roles in the outcome.

In addition, the available dataset (including just over 800 clients) is relatively small for modelling of this kind. We therefore could not expect our model to be as accurate as those developed in other sectors, where datasets are large and outcomes follow more predictable rules (such as models for determining whether an email is spam, or whether a cancer cell is malignant, etc.).

To evaluate the accuracy of our model, we split our dataset into two – using 80% of the data to train the model and holding back 20% for testing (7). In the end, when the model was tested against the unseen data, it performed moderately well at predicting client outcomes, but with a considerable margin for error.

Model training, overfitting, and unseen data

A machine learning model needs to be 'trained' on a dataset – this is how it learns which variables are most important for making its predictions. The process by which a model learns from a dataset is called 'fitting'. However, when it comes to testing a model, it is important to test it on new data that it was not trained on. This is because there is a danger of models becoming 'overfitted' – they become overly focused on the training data and pick up on tiny trends in it which are actually just 'noise' (i.e., the result of random chance that don't signify any meaningful relationships), meaning that when the model encounters new data its predictions are inaccurate. For this reason, a portion of data is always held back for testing. This is the 'unseen' data.

7 This process was conducted 1,000 times, with differing samples for the training and testing sets taken each time. The figures for the model's accuracy reflect the average performance over all 1,000 iterations.

In every performance metric, the model significantly outperformed the benchmark of a 'naïve' model (i.e., a model that guesses randomly). This tells us that the model does provide useful information (8). Crucially, the p value for the model was 5.636-27, giving us high confidence in the validity of the relationships it describes (9).

However, the proportion of errors made by the model when tested against unseen data remained quite high. The model had an accuracy of 74% (random guessing would yield an accuracy of 58%) (10). Similarly, the model had a precision of 54% (whereas random guessing would yield a precision of 28%) (11). As the model comfortably outperforms random guesswork, it is telling us about real trends in the data. However, it still made significant numbers of mistakes in its classifications.

This level of accuracy leaves a lot of room for error, meaning that the model's results need to come with several caveats. The model's predictions are far from perfect, and the estimates it produced should not be understood as perfectly accurate measurements. Instead, the findings are only our best estimates, in a field where uncertainty and unpredictability are unavoidable.

Nevertheless, we hope the findings can give a general impression of the underlying dynamics at play in the data. Where possible, we have included confidence intervals in our results to reflect this uncertainty, and to indicate the range of values within which we believe true figures are most likely to be situated.

8 A detailed breakdown of the model's performance is provided in Appendix 2.

9 A p value gives the probability that the correlations observed by the model occurred simply due to random chance, rather than due to real relationship between the variables. This is also known as the model's 'statistical significance'. Generally, p values below 0.05 are considered good (meaning there is less than a 5% chance that the relationships described in the model occurred due to random chance). Our model has a p value of 5.636-27, i.e., considerably smaller than 0.001.

10 This means that 74% of the model's predictions were correct. For an explanation as to why random guessing does not give an accuracy of 50%, see appendix 2.

11 This means that, when the model predicted a client would achieve employment, it was right 54% of the time. For an explanation as to why random guessing does not give a precision of 50%, see appendix 2.

Confidence intervals

Whenever an experiment is done repeatedly upon samples from the population, the results of the experiment will be slightly different each time. The experimenter would ideally like to know what the result would be if they ran the experiment on the whole population, but this isn't practically possible. Instead, they give a confidence interval to show how confident they are that the experiment they ran on the sample would accurately reflect the true result for the whole population (i.e. how reliable their experiment was). A confidence interval takes the form of a range of values and a level of confidence, and shows how confident the experimenter is that the true value is somewhere between the two figures. For example, if a 95% confidence interval of 5 and 10 is given, then the author is 95% confident that the true value for the population would be between 5 and 10.

A confidence interval is usually calculated statistically using the standard deviation of the results of the experiment (based on the inference that results from experiments using samples usually follow a consistent 'normal distribution' around the true value for the population). However, classification models often use a 'bootstrapped' confidence interval – this is more of a brute force solution, where the model is trialled hundreds of times on different samples of data, and the confidence interval indicates the range of values that 95% of the results fell between.

Meet a volunteer

My name is Sabrina and I've volunteered at Breaking Barriers as an Advice and Guidance volunteer since 2019. I am granddaughter to migrant grandparents who came to the UK for a better life. My surname is given by my father's Polish ancestry, my mother is a migrant born in Kenya with Indian heritage. My parents separated when I was a baby due to racism which has caused me to have strong feelings toward all forms of discrimination and racism. I believe everyone deserves choice and fair treatment anywhere in the world they go.

My passion for many years has been to mentor and coach people. I love to learn, and I believe in continuous growth to keep your skills fresh and your mind active. I have an extensive background working in corporate office environments in IT, HR, Insurance, Business Management, Project Management, and Talent Development. I now run my own recruitment agency supporting inclusive employers to hire diverse candidates.

I have enjoyed volunteering for multiple refugee charities since 2018, including Breaking Barriers. I find it very rewarding to help someone search and apply for jobs and create a CV. The job market is very competitive, and it can feel very overwhelming for those that don't speak English as a first language. I see the struggles people face and I love to do my best to guide others to prosperity and happiness.

Sabrina – Advice and guidance volunteer

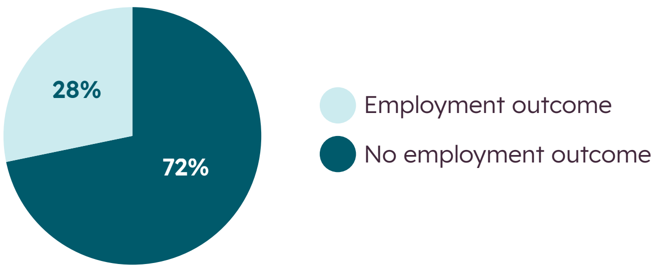
Summary of the dataset

The dataset used for this analysis contained all clients that enrolled with Breaking Barriers from financial year 2019 to financial year 2022 who were unemployed or inactive at the time of their enrolment (12). This generated a dataset of 825 clients. This section highlights some of the characteristics of these clients.

Just under one third of all these clients in the dataset entered employment during their time with Breaking Barriers (13).

Most clients completed under 6 hours of one-to-one support, with clients most frequently stopping before their fourth hour (at least before entering into employment – as hours completed after clients entered employment were excluded from the study). However, a small number of clients completed more than 20 hours.

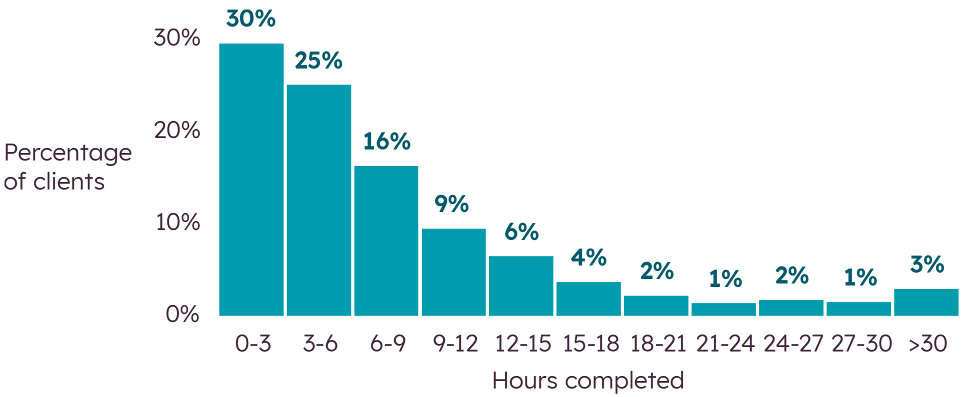
Clients achieving employment



12 Clients enrolled prior to financial year 2019 were excluded as insufficient client data was gathered at this point.

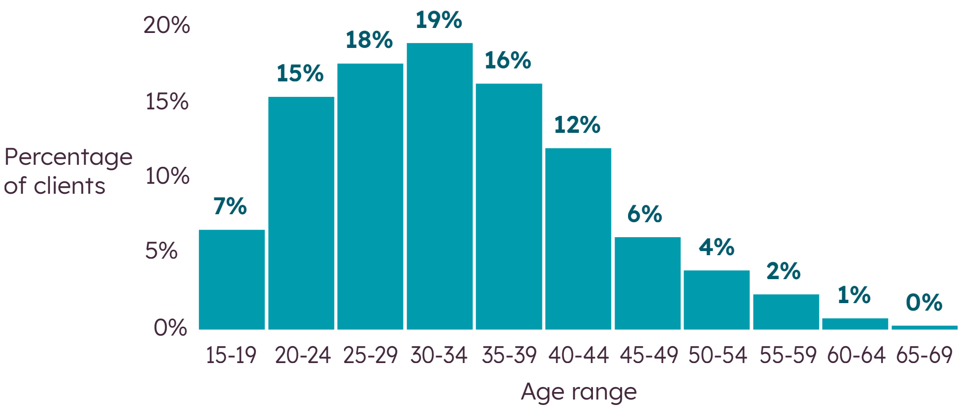
13 With many clients also achieving non-employment outcomes, such as entering education or training, in the same period.

Support hours completed per client

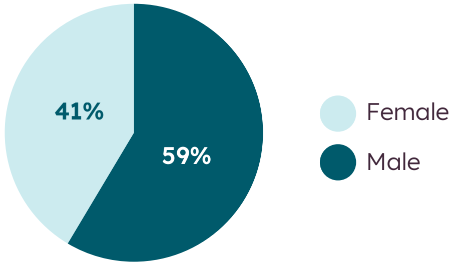


The majority of clients were under 40 years old, with clients most frequently enrolling between the ages of 25 and 35. In terms of gender, male clients were slightly more common than female.

Clients by their age at enrolement

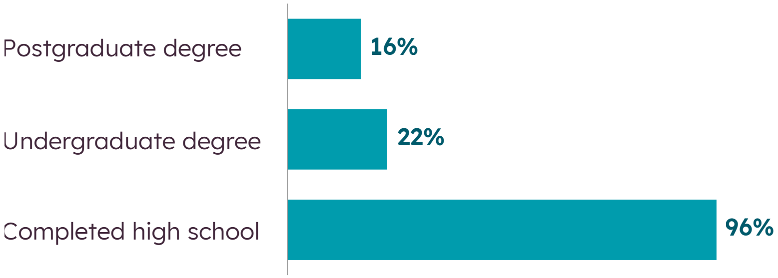


Clients by gender

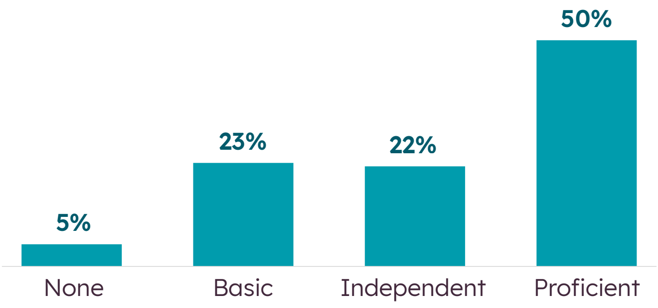


Almost all clients had completed high school and a smaller proportion had degree-level education or above. A majority of clients were proficient English speakers (using the CEFR English language rating system), and only one in twenty clients had no English knowledge at all.

Clients by level of education

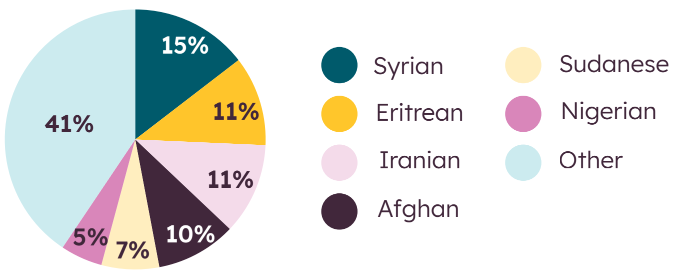


Clients by English language speaking ability



Clients came from 65 different countries, highlighting the incredible diversity within the client group. The most frequently occurring nationalities were Syrian, Eritrean, Iranian, Sudanese, Afghan and Nigerian.

Clients by country of origin

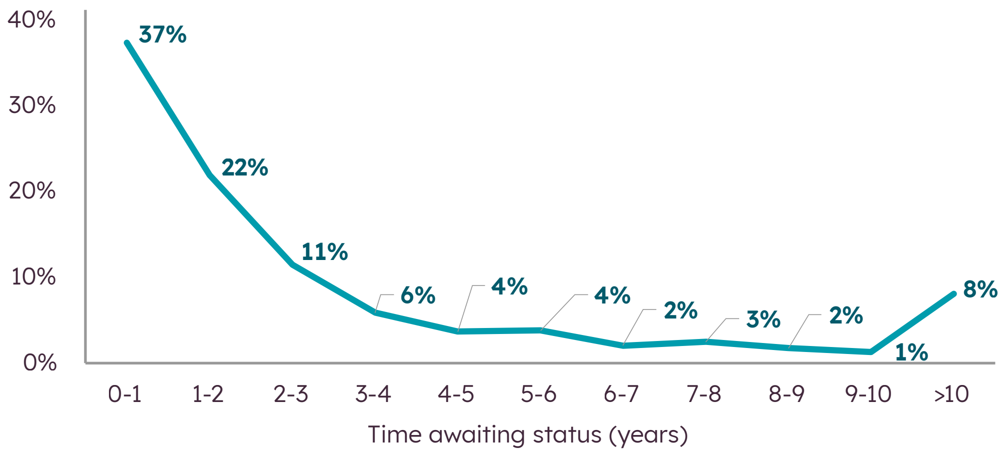


One relatively shocking feature of the clients included in our dataset was how long some had been in the UK awaiting their refugee status (while their asylum application was pending).

Although over a third had their applications approved within a year, almost one in ten were waiting over ten years for their status – an extraordinarily difficult situation considering that many asylum seekers are not allowed to work, and the small living allowances they are given.

Upsettingly, over a quarter (28%) of clients were formally homeless when they enrolled with Breaking Barriers. Around a third of clients (31%) cared for dependents at the time of enrolment. Most alarmingly, 7% of clients were both homeless and caring for dependents at the same time.

Time clients awaiting refugee status



Breaking Barriers' impact

The model estimated that our support sessions play a crucial role in increasing clients' chances of entering employment (14). All else being equal:

- the average client had an estimated probability of employment of 7% before they received any support (15).
- After 0-3 hours of support, this probability nearly doubled to 13% (16).
- After 3-8 hours of support, this probability more than tripled to 22%.
- If a client continued for 22 hours, their probability of employment reached 36%.

14 It should be noted that the p value for the relationship between support hours and probability of employment was significantly less than 0.001, so we can be confident in the existence of a relationship between support hours and employment.

15 These figures were calculated by re-running the model on hypothetical clients that had the mean values for all other variables but differed only in the number of support hours they had completed. The model then gave the difference in probability that could be attributed to each additional support hour.

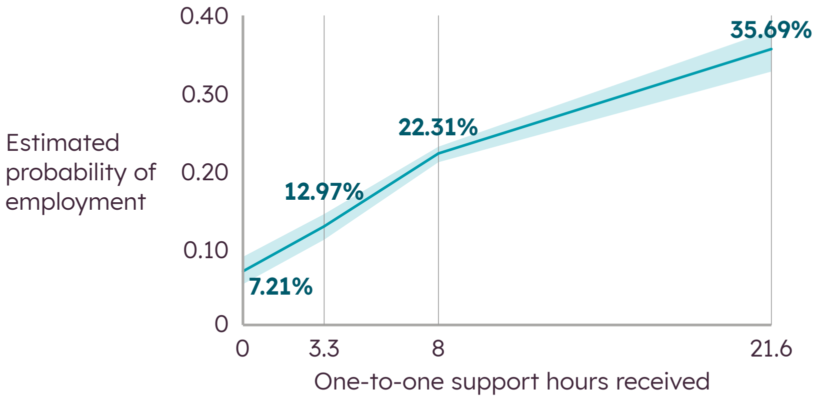
16 The support hours variable was transformed into a variable with 3 bins, as this was shown to maximise the model's accuracy. The bins were set at 0-3 hours, 3-8 hours, and 8-21 hours (based on the quantiles in the underlying data, with 33% of cases having 3 hours or less, 66% having 8 hours or less, etc.)

The graph shows the relationship between support hours and estimated probability of employment, with the shaded area illustrating a 95% confidence interval for the figures (17). It is notable that between hours 0 and 8, clients experience a rapid growth in their estimated probability of employment, which flattens subsequently. This could indicate that much of the value of our one-to-one sessions is delivered within the first 8 hours.

By running the model on an amended version of the dataset where all clients had 0 support hours, the model could be used to estimate what would have happened without our support. The model estimated that, in this scenario, 80% of the clients that entered employment would not have done so (with 95% confidence that the true figure is between 69% and 89%).

17 This is a "bootstrapped confidence interval", calculated by running the model 1000 times with a different sample of the data on each iteration. The values provided above represent the mean results. The lower estimate of the 95% confidence interval represents the results at the 0.025 quantile, and the upper estimate represents the results at the 0.975 quantile. The lower estimate values were 5% at 0 hours, 11% at 0-3 hours, 21% at 3-8 hours and 33% at 8-22 hours. The upper estimate values were 9% at 0 hours, 14% at 0-3 hours, 23% at 3-8 hours, and 38% at 8-22 hours.

One-to-one support hours and employability



Our model estimated that:

8 hours of one-to-one support sessions **more than tripled a client's chances** of finding employment (from 7% probability to 22%).

80% of all clients' employment outcomes **would not have occurred without the support they received** from Breaking Barriers.

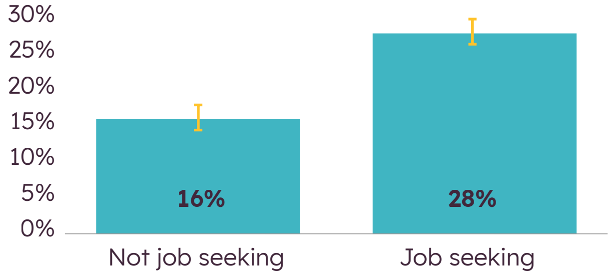
What else did we learn?

The model identified various factors that helped predict how likely a client was to find employment. Here are some of the key insights that the model highlighted in our data (18).

Prior experience job hunting

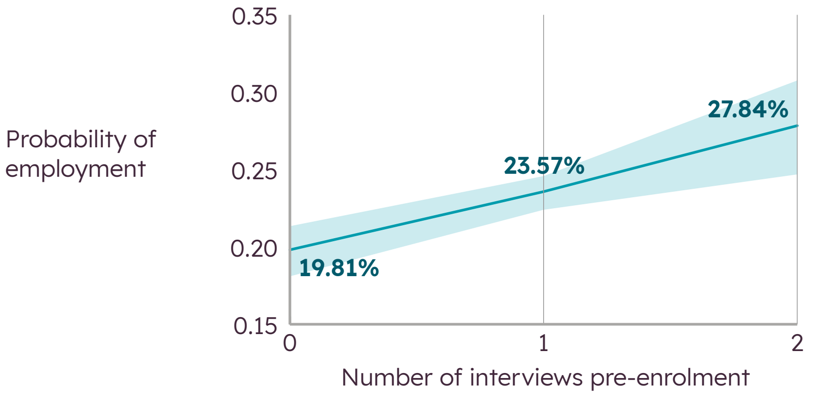
Unsurprisingly, a client's previous experience of job-hunting played a big role in how successful they were after enrolling with Breaking Barriers. Some clients had already been job hunting for some time when they enrolled with us, and so had more experience with putting together CVs and cover letters, whereas others had not yet started making applications. This experience made a big difference – those that were already job hunting had an estimated 28% likelihood of achieving employment, whereas those who had not had only 16% likelihood.

Client's probability of employment by whether they were already job-seeking prior to enrolment with Breaking Barriers



Equally unsurprisingly, clients' levels of success in their previous applications also made a difference to their later probability of entering employment. The number of interviews that clients had been offered prior to their enrolment with Breaking Barriers was positively correlated with their probability of entering employment.

Number of interviews and employability

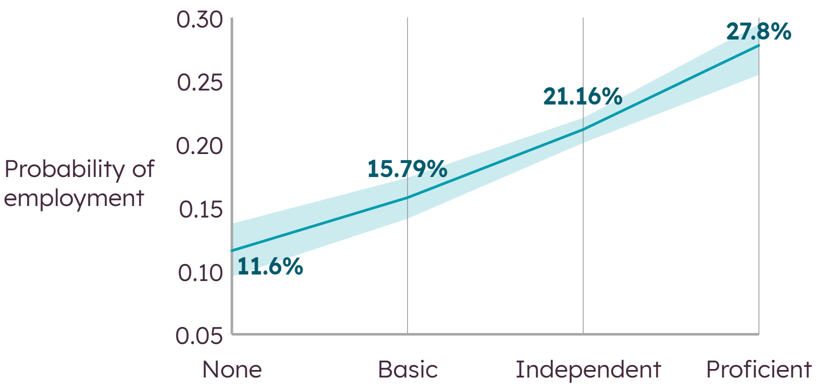


18 All figures in this section were calculated using the same method as in the previous section. By running the model on hypothetical clients with the mean values in every variable apart from one variable of interest at a time, we could isolate the model's estimations for the impact of specific variables (while controlling for the effects of all the others).

English language level

Another unsurprising finding was that client's English language ability made a big difference to their chances of finding a job. Clients' level of spoken English at enrolment (rated using the CEFR scale from Basic to Proficient) proved to be highly correlated with probability of employment. Clients with proficient English abilities had an estimated probability of employment of 28%, whereas those with basic abilities had a probability of only 16%.

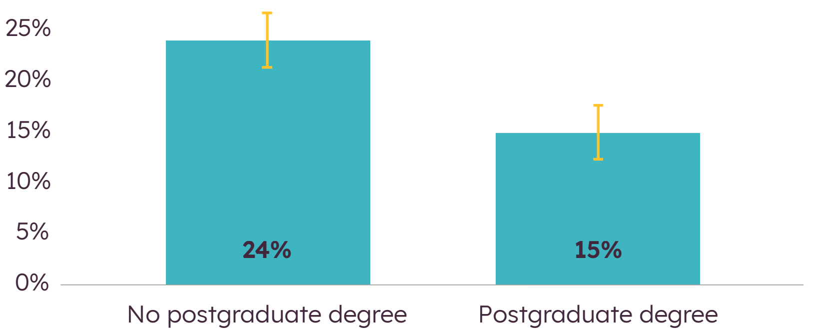
English speaking level



Level of education

Unexpectedly, we found that general levels of education had little to no effect on the model's estimated probabilities of employment. Whether clients had completed high school or had an undergraduate degree had statistically significant relationship with their employment rates. The exception to this rule was postgraduate education, however, the direction of the relationship was not what we might have expected. The model predicted that clients with postgraduate degrees were less likely to enter employment than those without. This might be because those with these high-level qualifications were only applying for much more competitive jobs, but the data cannot show this.

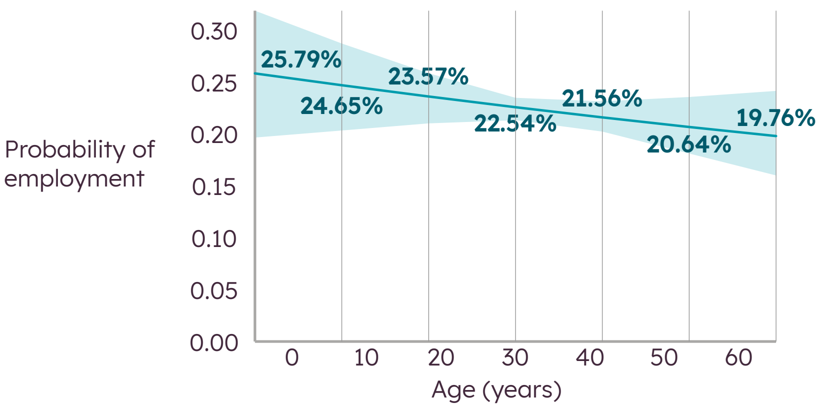
Probability of employment by whether the client has a postgraduate degree



Age

The model estimated that older clients were on average less successful than younger ones, falling from 25% at age 20 to 18% at age 60 (19).

Age at enrolment

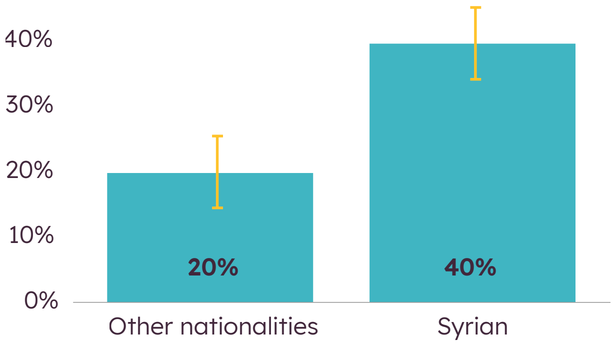


19 The graph shows the model's estimations for probability of employment in relation to age and extrapolated the relationship across a full range of ages – this should not be interpreted literally for ages below the age of 18 (the minimum age for Breaking Barriers clients), as of course a client aged 10 would not have a 27% chance of achieving employment!

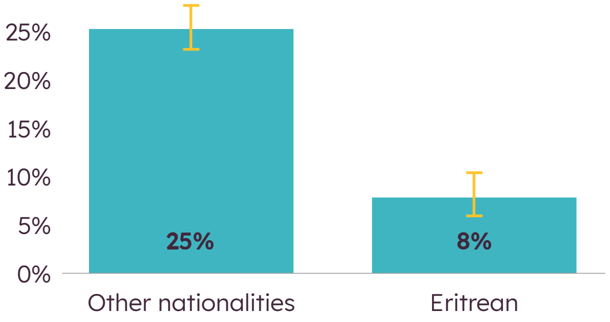
Nationality

Clients came from over 60 countries of origin, so it was impractical to test every nationality in a model. However, of the most frequently occurring countries of origin, two seemed to have significant impacts on the probability of a client finding employment. The model estimated that, all else being equal, clients from Syria were twice as likely as those from other places to enter employment, and clients from Eritrea were a third as likely as those from elsewhere. This is a sizeable effect, and one that, at present, we do not know how to explain, so requires further investigation.

Probability of employment by whether the client was Syrian



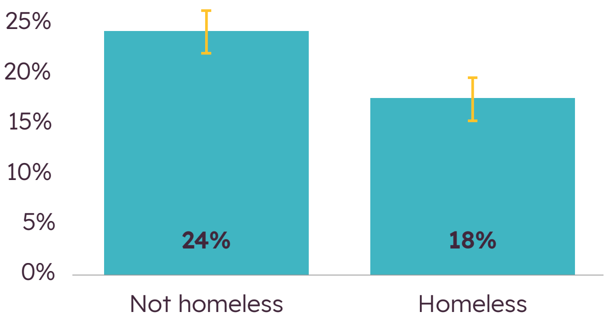
Probability of employment by whether the client was Eritrean



Homelessness

Clients that were formally homeless when they enrolled with Breaking Barriers were also, understandably, less likely to find employment (24% vs 18% estimated probability of employment).

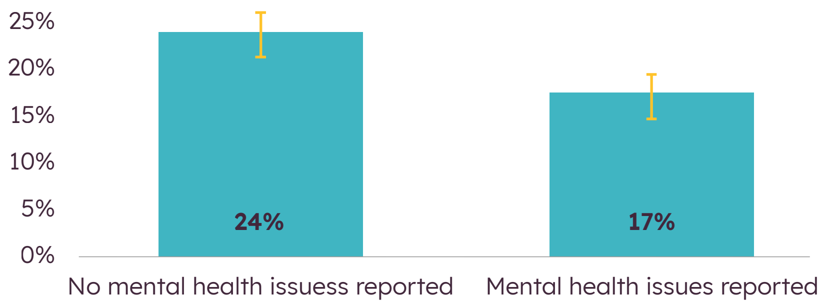
Probability of employment by homelessness status at enrolment



Mental health

The condition of a client's mental health can also play a key role in the model's estimation of their likelihood of entering employment, as mental health challenges can prevent clients from being able to put together applications or even start work once job offers are made.

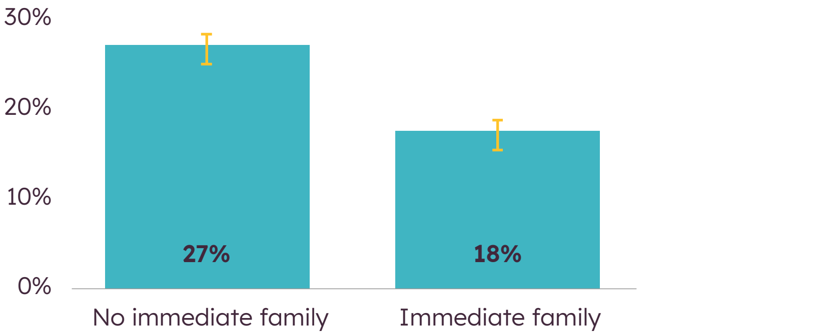
Probability of employment by whether the client reported mental health issues at enrolment



Immediate family in the UK

Whether a client had immediate family in the UK made a difference to their estimated probability of employment. However, the relationship was not in the direction we expected – we found that clients without immediate family in the UK were predicted to be more likely to enter employment. This could be picking up on other factors, such as caring responsibilities, or the other traits of clients that were more likely to have UK-based family (20). It might also be the case that these clients had other earners in their households and so were more financially able to hold out for better positions. However, we cannot explain the findings with certainty at present.

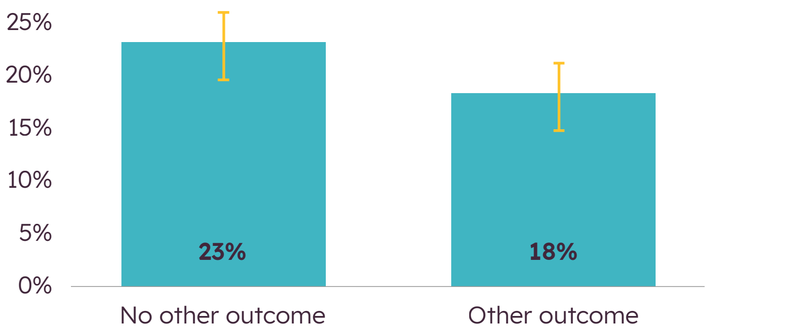
Probability of employment by whether the client has immediate family in the UK



Non-employment outcomes

As well as entering into employment, Breaking Barriers tracks clients achieving other kinds of outcomes – such as starting a qualification, doing work experience, or undertaking vocational training. Intriguingly, we found that clients who achieved these other non-employment outcomes were less likely to subsequently find employment. This does not, however, necessarily mean that those other outcomes were counter-productive – it might simply be the case that these outcomes might reduce someone's availability to work in the short term (for example, if they started a full-time degree course) but increased their employability in the long term.

Probability of employment by whether the client had already achieved a non-employment outcome

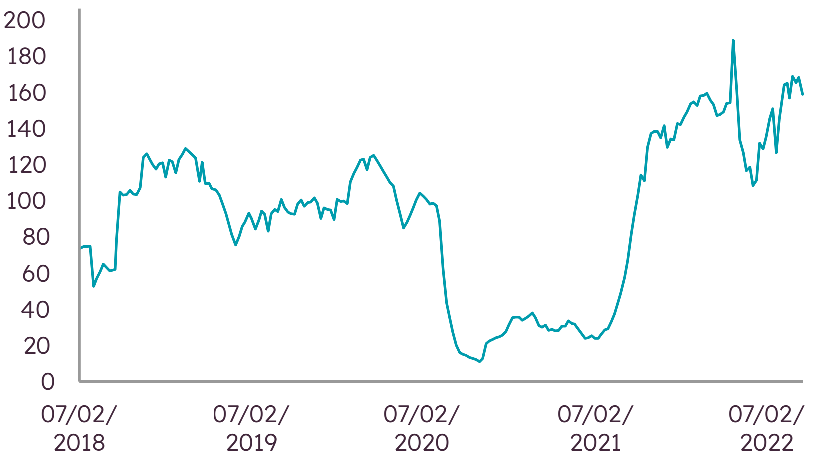


20 Although it should be noted that multi-collinearity tests were performed to ensure that no two variables were directly correlated with one another.

The job market

The condition of the labour market from financial year 2019 to financial year 2022 made a considerable difference to rates of success – particularly as the clients in our dataset were being supported during the various stages of the Covid pandemic. We incorporated labour market variables into our model using data on estimated job vacancies per week compiled by the ONS and Adzuna (21). The effects of the pandemic are clearly visible in the dataset:

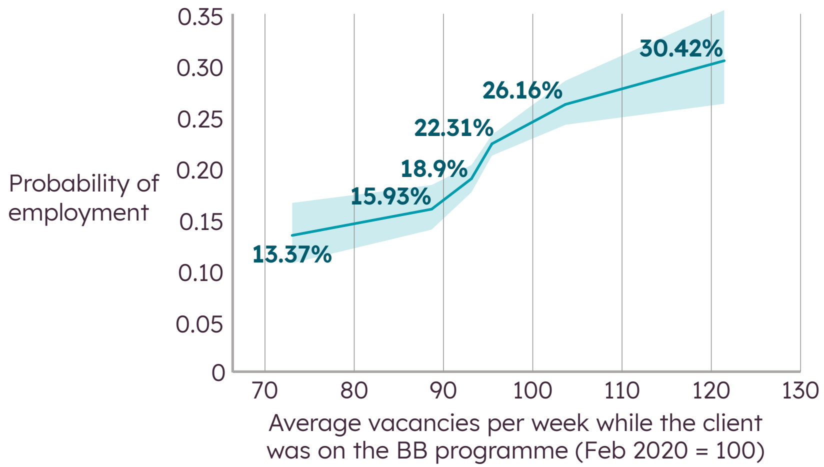
Hospitality job adverts per week (Feb 2020 = 100)



After trialing a variety of variables, we found that the number of job vacancies in the health and hospitality sectors during the period that clients were on the programme were statistically significant predictors of employment outcomes (both with p values below 0.05). However, the relationships were not quite as we expected.

Jobs in the health sector behaved more conventionally – the higher the number of job ads per week while a client was with Breaking Barriers, the more likely the client was to enter employment (22).

Health sector vacancies

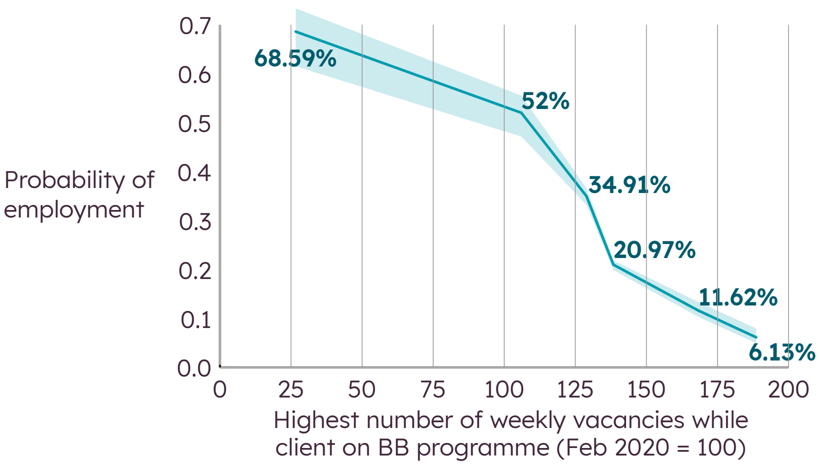


21 https://www.ons.gov.uk/economy/economicoutputandproductivity/output/datasets/onlinejobadvertestimates

22 Job ad figures provided by ONS are indexed to 100 at February 2020.

However, hospitality jobs displayed a much more unexpected relationship. We found that, the more jobs were available in the hospitality sector, the less likely a client was to enter employment (23). This relationship, while statistically significant, is highly counter-intuitive, and we do not yet know how to explain it (particularly as jobs in the hospitality sector were particularly sensitive to the effects of Covid lockdowns)

Health sector vacancies



23 For health sector job adverts, the best predictor was the average number of job ads per week while a client was on the programme. However, for hospitality job ads, the best predictor was the maximum number of jobs ads out of all the weeks the client was on the programme (i.e., in all the weeks that the client was being supported by Breaking Barriers, what was the highest value that the number of job ads per week reached).

Factors that did not influence probability of employment

Many additional variables were tested but found not to have any statistically significant influence on the estimated probability of employment in the model (24). These included:

- Gender
- Sexual identity
- Caring responsibilities

Some of these variables are quite surprising. For example, we would expect a client's level of education to influence their employability (although this perhaps reflects the extent to which UK employers discount qualifications earned overseas).

Similarly, we would expect clients with close relatives in the UK to have an advantage. These factors need further exploration to be fully understood.

24 This is not to say that we know definitively that these factors make no difference – it is simply that our data does not show any effects relating to these variables.

Conclusion

This research represents Breaking Barriers' first attempt at using statistical modelling to quantify its impact, as part of our commitment to being an evidence-based and data-led organisation.

The methodology used in this study comes with many caveats, and the accuracy of the model developed does leave plenty of room for error. However, in the absence of a randomised control trial, this approach still constitutes a significant step forward for our impact measurement.

Moreover, we can be encouraged by our results – with the model showing a clear relationship between one-to-one support hours and the probability of employment, and with the model estimating that 80% of clients' employment outcomes can be attributed to our support.

What next?

We are hopeful that we will be able to build on this research in future – improving the model by growing our dataset and gathering data on new variables. In future, modelling like this could be used to help us identify clients that need particularly intensive forms of support, enabling us to allocate our resources and design our programmes more effectively.

We could also deploy similar analytical methods to examine the programmes and client groups that were not included in this research (such as those clients that already had jobs upon enrolment).

Appendix 1: Building the model

Data was prepared for the model using the following steps:

1. Client data was filtered to only include clients enrolled from financial years 2019 to 2022 who were unemployed at enrolment. Categorical variables in the original dataset were recoded to Boolean and ordinal variables for analysis. Each client's number of support hours prior to entering employment were calculated.

2. Data for job ads was added to the client dataset (comprising calculations for the average number of job ads and maximum number per week of job ads while clients were being supported by Breaking Barriers.

3. Outliers were identified for each numerical variable (using a standard method, with an upper fence of 1.5 IQR above the third quartile and a lower fence of 1.5 IQR below the first quartile) and imputed with upper and lower fence values. Blank values were then imputed with means for each variable.

Variables were selected for the model using the following process (25):

1. Variables were checked for multicollinearity by running linear regression models of all variables against each other. If the regression model returned an R2 of greater than 0.4, the two variables were tested in single logistic regression models to assess their usefulness for predicting employment outcomes, and the variable with the lower p value in its logistic model was dropped from the dataset.

2. Variables were then checked to assess whether any transformations could be more predictive than the original variable (due to non-normal distributions in the data). Each numeric variable was tested against its exponent, log, and binned variables with 3, 4 and 5 bins (bin sizes were calculated using quantiles from the data, so for a 3-bin transformation the first bin was set to contain the lowest third of values, and the second bin to contain the next third, etc.). Each transformation was tested in a single logistic regression model, and the transformation with the lowest p value was included in the dataset. Notably, this led to one-to-one support hours being transformed to a 3-bin variable.

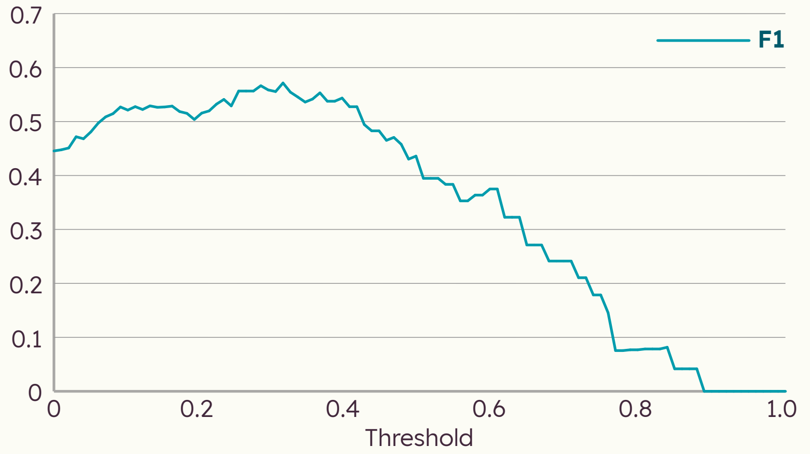
25 Logistic regression models were creating using the Statsmodels library in Python.

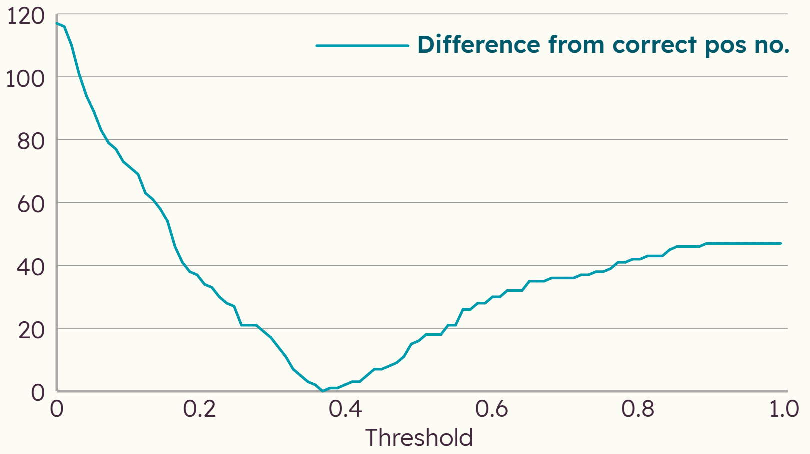
3. A test was performed to identify potential interaction variables. All variables were multiplied with each other, and the predictive power of the resulting interaction variables was tested in single logistic regression models. To be included in the dataset, an interaction variable had to have a p value in its logistic model of less than 0.4 and to have a p value that was at least 0.1 below the lower of the two individual p values of the variables that constituted the interaction. In addition, for Boolean interaction variables, a meaningful number of clients had to have each value for the variable to be included. Ultimately, no interaction terms met these criteria and none were included in the model.

4. A first multiple logistic regression model was created using all the variables in the dataset resulting from the above steps. This provided a breakdown of p values for each variable, controlling for all others. Any variables with a p value over 0.3 were then removed.

5. The resulting model was then run with different thresholds for positive classification, ranging from 0.01 to 0.99. The threshold of 0.37 was selected for the final model for an optimisation of F1 score and the accuracy of the number of positive classifications it made (see charts below).

6. The final model was run 1000 times with different samples from the dataset (using the random state parameter) to get average values and bootstrapped confidence intervals for key figures such as F1 score, precision, accuracy, and recall, as well as the estimated attribution rate.





Variables included in final model:

Variable name	**Data type**	**P value**
Constant	Numerical	0.001872
Hours of one-to-one support (3 bins)	Numerical	0.000001
Non-employment outcome	Boolean	0.098916
Age at enrolment	Numerical	0.215075
Ethnicity = Black	Boolean	0.063132
Ethnicity = Asian	Boolean	0.279266
Homelessness	Boolean	0.115947
Mental health issues	Boolean	0.040207
Postgraduate degree	Boolean	0.0094
Immediate family in the UK	Boolean	0.036243
Syrian	Boolean	0.000683
Eritrean	Boolean	0.007452
Job seeking prior to Breaking Barriers	Boolean	0.001982
English language speaking level	Ordinal	0.001951
Number of interviews before enrolment	Numeric	0.065242
Hospitality sector maximum job ads (5 bins)	Numeric	0.000001
Health sector average job ads (5 bins)	Numeric	0.06104

Appendix 2: Detail on model evaluation

The table below provides a detailed breakdown of the model's performance against unseen data using various metrics. The model's performance is compared to that of a benchmark of a 'naïve' model (that guesses randomly).

It should be noted that the random guesswork of the benchmark naïve model does not yield an accuracy of 50% because only 28% of clients in the dataset did successfully enter employment. As the data is skewed in this way, the benchmark naïve model used 'stratified' random guesses (i.e., it randomly allocated 28% of clients as achieving employment and 72% as not achieving employment, rather than allocating 50/50). For calculating the naïve model log loss, every client was given a probability of 0.28.

Measure	Description	Score
Accuracy	Proportion of predictions that were correct. Random guessing would give an accuracy of 59%.	74%
Precision	Proportion of predicted employment outcomes that were correct (true positives/all positive predictions). Random guessing would give a precision of 28%.	54%
Recall	A proportion of actual employment outcomes identified by the model (true positives/actual positives). Random guessing would give a recall of 28%.	57%
F1 score	Harmonic average of precision and recall. This is one of the most widely relied on measures of model reliability. Random guessing would give an F1 score of 28%.	0.55
McFadden's Pseudo R2	Proportion of variability in the probability of employment that can be explained by the model. A McFadden's Pseudo R2 between 0.2 and 0.4 is considered a very good model.	0.22
Log loss	A measurement of the average error in estimated probabilities. Random guessing gives a log loss of 0.60.	0.50
P value	Likelihood that the effects identified by the model are erroneous and occurred by chance. Below 0.05 is considered 'statistically significant'.	$6.535e^{-28}$

Breaking Barriers

Meaningful employment for refugees

Thank you for reading this report. If you have any questions about the data in this report please contact Toby Gill at: t.gill@breaking-barriers.co.uk

breaking-barriers.co.uk

enquiries@breaking-barriers.co.uk

/breaking-barriers @bb_uk1 @breakingbarriersuk