

Research Bulletin 18/10 | Factors Influencing Positive Outcomes on the Training for Success Programme

Carole-Ann McKay, Nicole McKee and Jamie Stainer, Analytical Services, Department for the Economy

December 2018

Summary

Data science techniques can be used to gain meaningful insights from large amounts of complex data and these insights can be used to make better decisions. This article presents the results of a Data Science project to identify factors which influence the likelihood of an individual having a successful outcome on the Training for Success (TfS) programme.

The results suggest that factors which influence success include the TfS supplier, individual's qualifications on entry, socio-economic status and whether a participant has a disability. The research demonstrates the potential value that can be added through the use of Data Science techniques and the new insights that predictive analytics can provide, to inform decision making and improve outcomes.

Introduction

What is Data Science and how can it help improve service delivery?

Data Science is defined as "applying the tools, methods and practices of the digital and data age to create new understanding and improve decision-making". It involves using new tools and innovative methods to provide rich, informed measurement and analyses on the economy, the global environment and wider society. Data Science can help to address specific research or policy-based questions by analysing data in new ways, making use of new technology, software and modelling techniques.

In the context of this research, predictive modelling allows service providers to identify those individuals who are at risk of non-achievement on the TfS programme at the point of enrolment. This information will allow providers to offer additional support to 'at risk' young people in order to help them achieve their potential.

Statisticians from Analytical Services in DfE were invited to partner with the Data Science Campus at the Office for National Statistics (ONS)ⁱⁱ to deliver the project, under the mentorship of an ONS Data Science expert as part of an initiative to build Data Science capability across the NICS.

Background to Training for Success

'Education, skills and employability' is one of the five pillars for growth in the draft Industrial Strategy for Northern Ireland (NI)ⁱⁱⁱ. The Department for the Economy (DfE) aims to "equip all of our people with the education and skills that will allow them to achieve their full potential" and to "support those who need our help to access the skills that will help them compete for employment opportunities". These aims are linked to the NI draft Programme for Government^{iv} outcomes, for example - Outcome 14, 'We give our children and young people the best start in life'. As part of achieving this ambition, the Department provides training to young people via the TFS programme.

Training for Success is a programme designed for young people aged 16-17 (or under 22/24 years for those with extended eligibility entitlement) which provides training to give them the tools and skills they need to get a job. The training provides young people with relevant qualifications as well as the required personal and behavioural skills to progress into work. Achievement rates^v on the TFS programme are consistently lower than agreed targets and improving these rates is the key to helping more young people to achieve the skills and qualifications they need to progress into Further Education, an apprenticeship, or employment.

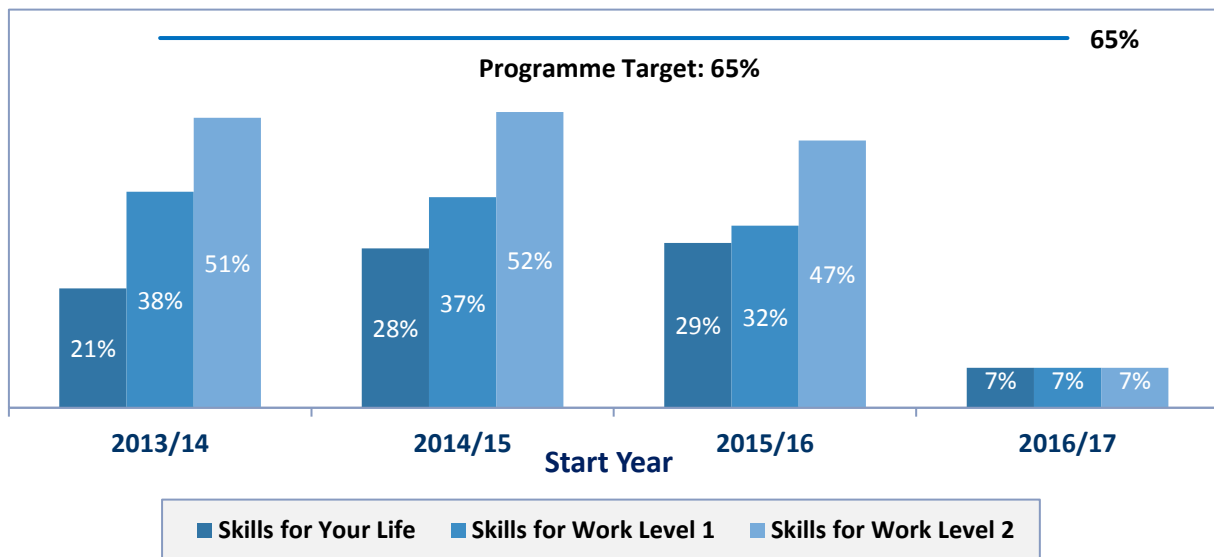
Framing the Problem

In this research, the team wanted to understand the key determinants of success on the TFS programme and address the question 'What factors influence a successful outcome?' This Research Bulletin Article presents brief descriptive analyses of current success on the programme and the results of a logistic regression based exercise designed to show which factors are stronger determinants of success.

The TFS programme is delivered across 4 strands: Skills for your Life and Skills for Work levels 1-3^{vi}. According to the most recently available statistics, there are 4,117 participants currently on Training for Success 2013/2017. The majority of participants on Training for Success 2013/2017 are male, accounting for 72% of current occupants. Over the duration of the programme to date, 59% of Skills for Your Life and 65% of Skills for Work leavers achieve a qualification^{vii}.

Achievement rates on the TFS programme are generally low, below agreed programme targets, although it should be noted that the agreed programme targets are aspirational and not based on historic achievement data. Understanding what factors affect the likelihood of an individual having a positive outcome on TFS would help policy makers to better understand performance of the programme as a whole and identify areas where improvements could be made to deliver a better service to young people. Figure 1 shows the performance of the different TFS strands against the agreed programme target of 65% achieving 'all targeted qualifications'. Due to the small number of participants, figures for Skills for Work Level 3 are not shown. While the success rate of people achieving all their targeted qualifications is low, the majority of TFS participants do gain at least 1 qualification.

Figure 1: Performance by Strand against agreed programme target for achieving 'All Targeted Qualifications'



The target year for achievement is 2 years after starting the programme so the majority of the 2016/17 cohort would not have completed their qualification at the last data extract which was April 2018.

Insights

The analysis was focused on the achievement of any qualification on the TfS programme, rather than achieving all targeted qualifications. This gave a larger sample to work with and a higher number of 'success' cases to build predictive models on, making the models more robust. A representative sample of TfS participants was drawn from across a number of academic years and the factors influencing success were explored. The sample size was 14,675 participants, 66% were male, 34% female. Of the total sample, 60% gained a qualification, 40% did not gain a qualification. This is broadly in line with the official statistics for TfS, with 59% of Skills for Your Life participants gaining a qualification over the duration of the programme to date and 65% of Skills for Work participants gaining a qualification over the same period.

Success rates in the sample data were analysed across a number of different factors. There was little variance in success rates across most factors, perhaps due to the fact the TfS client group is quite homogenous and there is not a wide variation in the demographic make-up of the TfS population, for example, the programme is age-limited. The success rates of some of the factors analysed are shown in the figures below.

Figure 2 shows the success rate by gender. Success rates between males and females are similar, 60% for males and 59% for females. It should be noted that males are more heavily represented in the sample (66%) and on the TfS programme in general (72%).

Figure 2: Success by Gender on Training for Success Programme.

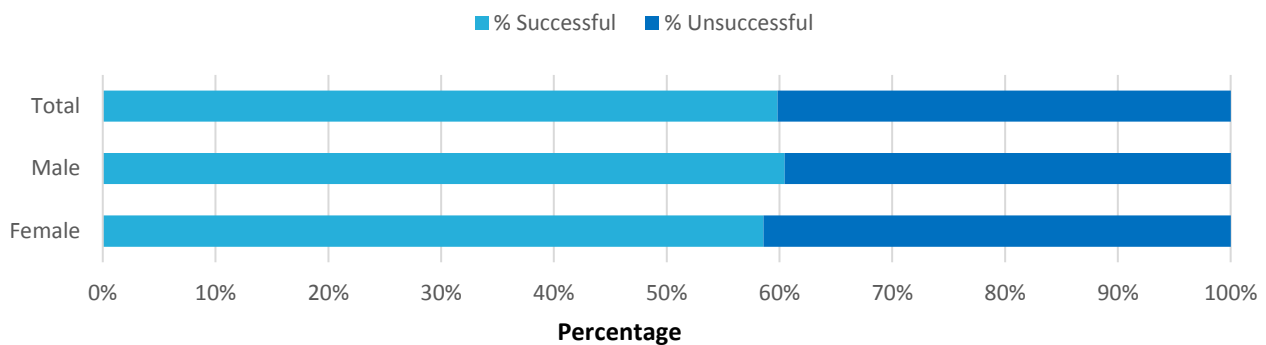


Figure 3 shows the success rate by TfS strand. Across the programme strands, Skills for Your Life had the lowest success rate at 55%, Skills for Work Level 1 had a success rate of 60% and Skills for Work levels 2 & 3 had the highest success rate of 73%. These 2 levels were combined due to the very small number of participants at level 3. It is unsurprising that Skills for Work Levels 2 & 3 had the highest success rate, as participants on these strands will, in most cases, have higher prior educational achievement than those in the Entry Level or Level 1 strands.

Figure 3: Success by Strand on Training for Success Programme.

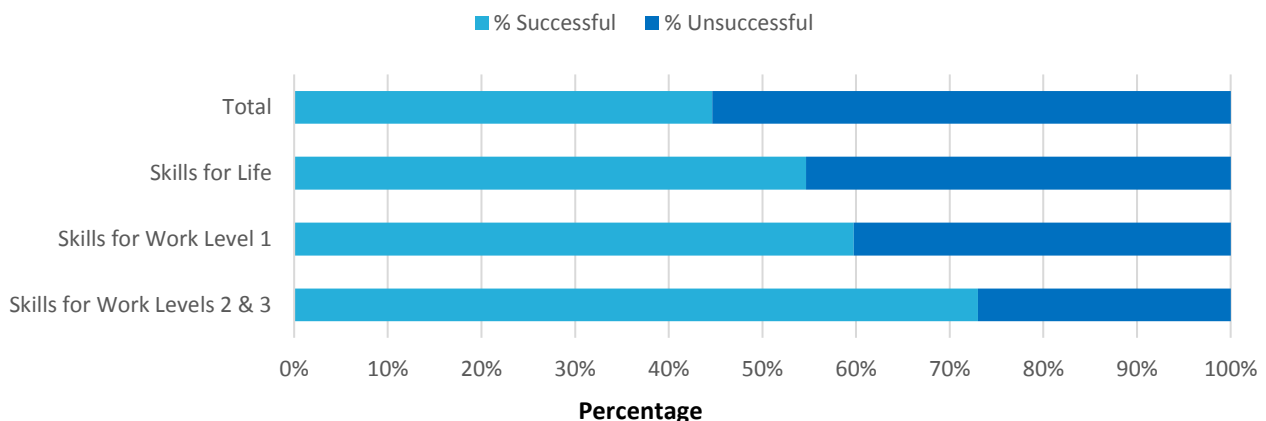
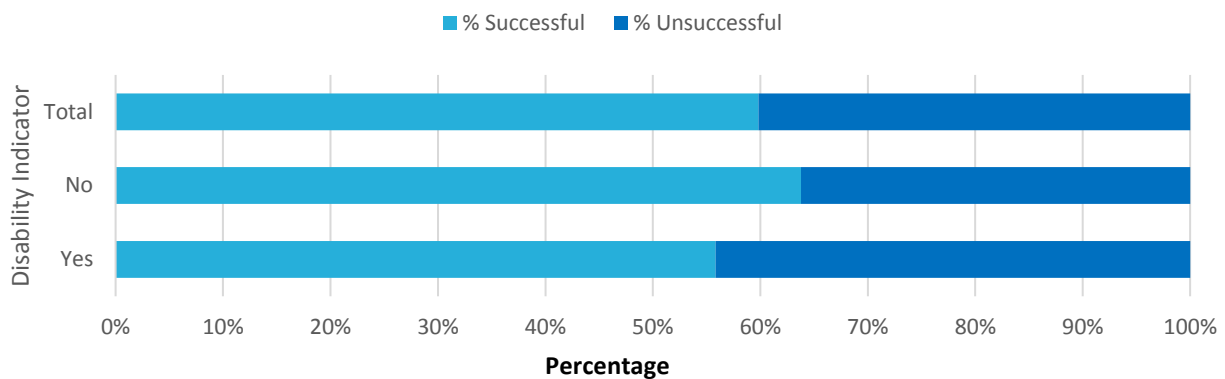


Figure 4 shows the success rates by disability indicator. Individuals without a disability had a higher success rate at 64%, while those with a disability had a success rate of 56%. It should be noted that the disability indicator on TfS is based on evidence of the young person having a disability as defined by the Disability Discrimination Act 1995 (DDA) and the Special Educational Needs and Disability (NI) Order 2005 (SEND0) and 50% of participants indicated that they had a disability. Individuals with a disability are entitled to extended eligibility and additional time and support on the programme.

Figure 4: Success by Disability Indicator on Training for Success Programme.



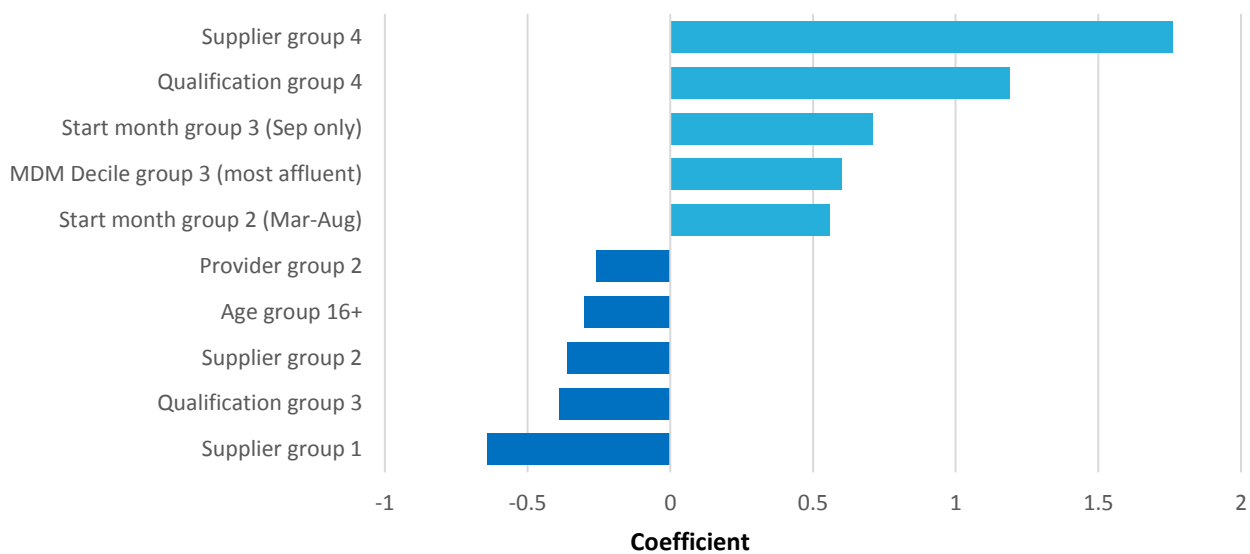
Regression Analysis

From the analysis of the TfS data above, it's possible to make some broad comments on the type of factors that may make an individual more likely to succeed on the programme. However, to properly assess these relationships a more robust analysis is needed. A number of different modelling techniques were used in order to build the best model possible to explain the factors that affect the likelihood of a positive outcome. This section of the research bulletin will present the results of a stepwise logistic regression model that had an accuracy score of 68%, i.e. it predicted the correct outcome for an individual 68% of the time.

The dependent variable in the model is 'success' – whether an individual achieved any qualification on the TfS programme or not. The independent variables included age^{viii}, gender, TfS supplier, disability indicator, qualifications on entry, Local Government District in which the participant is resident, start month and socio-economic status, using the NI multiple deprivation measure as a proxy^{ix}. The categories within some of these variables were grouped together to allow for analysis. So, for example, the supplier variable which contains over 60 different suppliers, was condensed into 4 supplier groups. Individual suppliers cannot be identified due to commercial sensitivities.

Factors influencing the likelihood of success can either have a positive effect, making success more likely, or a negative effect, making success less likely. Most of the results of the analysis are statistically significant at the one percent level. Figure 5 below shows the top 5 positive and top 5 negative factors in the model. The greater the coefficient score, the more impact the factor had on likelihood of success.

Figure 5: Results of Logistic Regression Analysis – Top 5 Positive and Top 5 Negative Factors



From these results, it can be seen that the factors with the largest effects in the model are Supplier and the individual's qualifications on entry. These variables can have both positive and negative effects on the likelihood of success. Other important factors include deprivation, with those in the most affluent areas having a higher likelihood of success and age, with those who are aged over 16 having a lower likelihood of success. Other factors which emerged as having a statistically significant impact that are not shown above include disability indicator; TfS strand, which is correlated with prior educational achievement; and some Local Government Districts, which suggests there may be some geographical or area-based effects.

Start month emerged as a significant factor in the model, with those individuals who started the programme in September having a higher likelihood of success than other participants (See Table 1). As the TfS programme is designed to be accessed at any time, this result seems unusual and could warrant further investigation. Analysis of the data suggests that individuals who start on the programme in September have a success rate of 64%, while participants who start in all other months have a combined success rate of 49%. It should be noted that 73% of all participants start their TfS programme in September. This variation in success rates might be explained by differences in the client group who start in September versus all other months, for example, those starting in other months may be more likely to have previously dropped out and be re-enrolling on the programme.

Table 1: Number of Starts and Success rate by Start Month

Start Month	Total Starts	% Successful	% Unsuccessful
September	10,694	64%	36%
All other months	3,981	49%	51%
Total	14,675	60%	40%

Discussion

The results of the logistic regression present some interesting findings. In particular, the difference in success rates of people who start in September versus other months may be of interest, as the programme is designed to be accessed at any time.

The most important factor in the model was Supplier. This may be due to differences in the type of client group that access TfS from different suppliers, or variance in the success rates of different professional and technical qualifications offered on different pathways. There may also be variations in the quality of training delivery, however the Education Training Inspectorate (ETI)^x carry out regular inspections of TfS programmes to ensure the quality of training provided is high and consistent across the TfS programme.

A limitation of the research was the quality of the administrative data used for the analysis. A number of variables had to be discarded due to incomplete or inconsistent data capture. A lot of the information is self-reported and data which are not directly tied to receipt of funding are not subject to the same strict audit controls as other data items required to process payments to suppliers. In addition, recording can vary across different suppliers and providers. The data are primarily recorded for operational purposes rather than for statistical analysis. However, a new recording and reporting system that is currently under development may address many of these issues.

Benefits of the Data Science Approach

The research team explored a number of other modelling techniques and algorithms, for example Random Forest^{xi} modelling and Adaptive Boosting^{xii}. The small size of the dataset meant it was not ideal for machine learning techniques such as these, but the results of these models were broadly similar to the logistic regression results, allowing a high degree of confidence in the reliability of the results, as they can be replicated across models.

The Data Science approach allows for new technologies, software and techniques to be used to generate new insights across research and policy making. Innovative work in this area is already being done across all sectors of the economy, including the public sector, for example in England and Wales. Data Science techniques allow for actionable insights into data to be produced. The techniques explored in this research have the potential to be used for operational purposes at a future point. For example, risk scores for an individual's likelihood of success could be generated at the point they enter the programme. This would allow service providers to identify individuals at risk of

not achieving, in order to target interventions and resources at them more effectively, potentially improving success rates and outcomes for young people.

Using predictive analytics can help the department and wider NICS deliver services more effectively, by providing new insights and a robust evidence base for decision making, while new software and modelling techniques can add value to existing data sources by interrogating and analysing them in new ways. The insights gained from these new methods have the potential to be operationalised and help to target resources to those most in need of support and to deliver better services to customers which will ultimately lead to better outcomes for the wider population.

Carole-Ann McKay, Nicole McKee and Jamie Stainer

For further information or queries please contact carole-ann.mckay@economy-ni.gov.uk

ii The ONS Data Science Campus was established in 2017 with the aim of building data science capability across the UK and beyond. <http://datasciencecampus.ons.gov.uk>

iii Northern Ireland Draft Industrial Strategy: <https://www.economy-ni.gov.uk/sites/default/files/consultations/economy/industrial-strategy-ni-consultation-document.pdf>

iv Northern Ireland Draft Programme for Government: <https://www.northernireland.gov.uk/programme-government>

v Training for Success Statistical Bulletins: <https://www.economy-ni.gov.uk/articles/training-success-statistics>

vi Skills for your Life is designed to address the personal and development needs of young people who have disengaged from learning and/or have significant barriers to education, training or employment. Skills for Work comprises a common curriculum of Essential Skills, personal and social development skills, employability skills and professional and technical skills, from levels 1 to 3, helping young people to progress into Further Education, Employment, or an Apprenticeship.

vii Training for Success Quarterly Statistics from May 2013 to April 2018, published August 2018 <https://www.economy-ni.gov.uk/sites/default/files/publications/economy/Training-for-Success-Statistical-Bulletin-Aug-18.pdf>

viii For the purpose of this model, the age variable was split into two groups. Group 1 was TfS participants aged 16 only, group 2 was all other TfS participants (aged 17-23, including extended eligibility for those with disabilities).

ix Northern Ireland Multiple Deprivation Measure: <https://www.nisra.gov.uk/statistics/deprivation>

x Education Training Inspectorate: <https://www.etini.gov.uk/>

xi Random Forest modelling overview: <http://www.statsoft.com/textbook/random-forest>

xii Adaptive Boost overview: <https://machinelearningmastery.com/boosting-and-adaboost-for-machine-learning/>