

# Department for Education

## How do Key Stages matter for child outcomes?

Lucy Ross, Politics and Quantitative Methods

### Objectives

The aim of my Q-Step project was to use longitudinal data to better understand the role of disadvantage in academic trajectories through the Key Stages (KS) of education.

Upon starting my internship, I set three main personal goals with my supervisor to achieve over the summer:

1. To develop my skills in SQL Server (a Microsoft database management system) and use this to prepare and clean my dataset for analysis.
2. To develop my skills in R to analyse the dataset and ultimately produce a policy-relevant report detailing the results of my analysis.
3. To gain a stronger understanding of the work of the department and the Civil Service more broadly.

### Methods

To create the longitudinal dataset for analysis, I used data from the National Pupil Database and matched this to an additional dataset to strengthen the analysis. I used SQL Server to match the data, and I then cleaned the data and created composite variables in SQL Server. My final dataset for analysis comprised of more than four million rows, spanning seven cohorts of learners.

I carried out analysis for KS2, KS3 and KS4 using 'achievement groups'. I created the groups in SQL Server based around score distribution at each KS. These allowed me to track movement meaningfully and consistently throughout different key stages and cohorts.

I used descriptive statistics first to identify the most influential variables at the different key stages of education and the way the variables interacted. I then developed histograms, stacked bar charts and boxplots using R and Excel to illustrate my analysis (see Figure 1 for an illustrative example).

I then used regression modelling to predict academic outcomes at each KS. I then used these predictions to create trajectory maps for key intersectional groups, illustrating the models' outcomes (see Figure 2 for an illustrative example). I carried out the regression in R and produced trajectory maps in Excel.

I used multiple linear regression for KS2, KS3 and KS4 and binary logistic regression for Further Education and Higher Education.

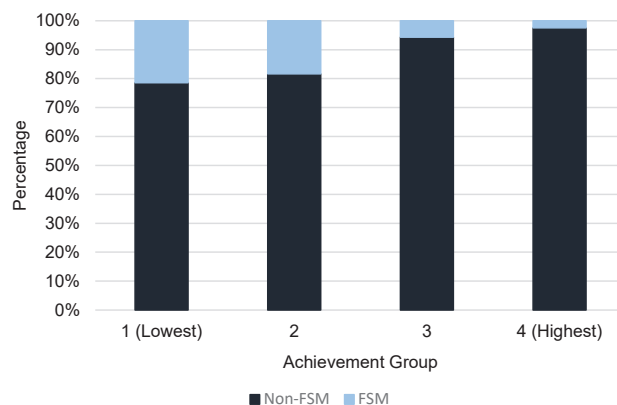


Figure 1: Example stacked bar chart illustrating the difference in achievement group compositions based on children's Free School Meal status (for illustration only, not based on actual data)

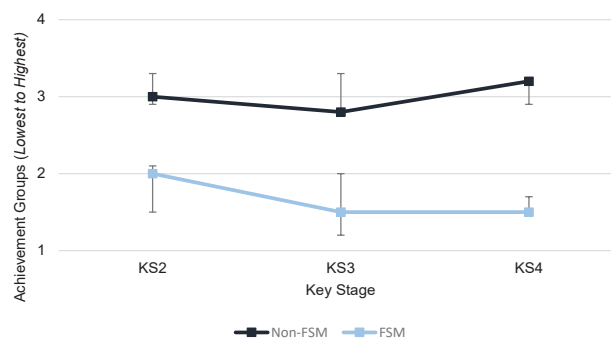


Figure 2: Example trajectory graph illustrating the different achievement group trajectories of children in and not in receipt of Free School Meals (for illustration only, not based on actual data)

### Outcomes

As the data used for my Q-Step project are currently unpublished, the complete findings of my research cannot be disclosed at this time. However, much of the work supported existing findings such as the impact of being in receipt of Free School Meals (FSM).

In terms of personal goals:

1. I feel competent in SQL and was able to successfully create, match and clean datasets.
2. I feel competent in R and was able to produce descriptive statistics, visuals and regression models using code, as well produce a policy-relevant report and presentation at the end of the internship.
3. I took the opportunity to meet with many civil servants both within the department and in other departments which really developed my understanding of the Civil Service.